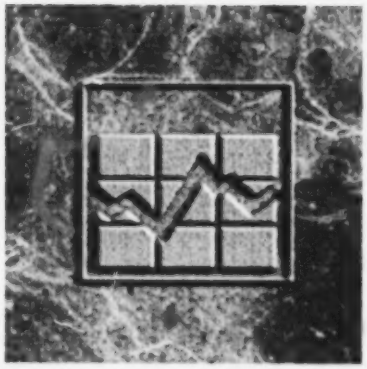
Catalogue no. 62F0026MIE — No. 003
ISSN: 1708-8879
ISBN: 0-662-40953-1

Research Paper

Household expenditures research paper series

2001 Food Expenditure Survey
Data quality indicators



by: Sylvie Auger, Khushnood Khan, Sylvain Nadon and Johanne Tremblay

Household Survey Methods Division
R.H. Coats Building, Ottawa, K1A 0T6

Telephone: 613 951-7355

Statistics Canada Statistique Canada

Canada

How to obtain more information

Specific inquiries about this product and related statistics or services should be directed to: Income Statistics Division, Statistics Canada, Ottawa, Ontario, K1A 0T6 (telephone: (613) 951-7355); (888) 297-7355 income@statcan.ca.

For information on the wide range of data available from Statistics Canada, you can contact us by calling one of our toll-free numbers. You can also contact us by e-mail or by visiting our website.

National inquiries line	1 800 263-1136
National telecommunications device for the hearing impaired	1 800 363-7629
Depository Services Program inquiries	1 800 700-1033
Fax line for Depository Services Program	1 800 889-9734
E-mail inquiries	infostats@statcan.ca
Website	www.statcan.ca

Information to access the product

This product, catalogue no. 62F0026MIE, is available for free. To obtain a single issue, visit our website at www.statcan.ca and select Our Products and Services.

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner and in the official language of their choice. To this end, the Agency has developed standards of service that its employees observe in serving its clients. To obtain a copy of these service standards, please contact Statistics Canada toll free at 1 800 263-1136. The service standards are also published on www.statcan.ca under About Statistics Canada > Providing services to Canadians.

Statistics Canada
Income Statistics Division

Household expenditures research paper series

2001 Food Expenditure Survey Data quality indicators

Published by authority of the Minister responsible for Statistics Canada

© Minister of Industry, 2005

All rights reserved. The content of this publication may be reproduced, in whole or in part, and by any means, without further permission from Statistics Canada, subject to the following conditions: that it is done solely for the purposes of private study, research, criticism, review, newspaper summary, and/or for non-commercial purposes; and that Statistics Canada be fully acknowledged as follows: Source (or "Adapted from", if appropriate): Statistics Canada, name of product, catalogue, volume and issue numbers, reference period and page(s). Otherwise, no part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopy, for any purposes, without the prior written permission of Licensing Services, Marketing Division, Statistics Canada, Ottawa, Ontario, Canada K1A 0T6.

July 2005

Catalogue no. 62F0026MIE2005003

ISSN: 1708-8879
ISBN: 0-662-40953-1

Frequency: Occasional

Ottawa

La version française de cette publication est disponible sur demande (nº 62F0026MIF au catalogue).

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued cooperation and goodwill.

Abstract

The Food Expenditure Survey (FES) is a periodic survey collecting data from households on food spending habits. Data are collected mainly using weekly diaries of purchases that the respondents must fill in daily during two consecutive weeks.

The FES, like all surveys, is subject to error despite all the precautions taken at the various stages of the survey to control them. Although there is no exhaustive measure of a survey's data quality, certain quality measures taken at various stages of the survey can provide the user with relevant information to ensure sound data interpretation.

This paper presents, for the 2001 FES, the following quality indicators the coefficients of variation, the non-response rates, the vacancy rates, the slippage rates, the imputation rates as well the impacts of imputation on the estimates.

Table of contents

Highlights 6

Introduction 8

1. Sampling errors 9

1.1 Measurement of sampling error 9
1.2 Coefficients of variation 9
1.3 Model for deriving an approximation of the CV 11
1.4 Suppression of unreliable data in tables of estimates 12

2. Non-response 13

2.1 Non-response rates and vacancy rates 14
2.2 Non-response according to the level of urbanization 16
2.3 Non-response by income strata 18
2.4 Adjustment for non-response 19

3. Coverage errors 19

3.1 Undercoverage and overcoverage: slippage rates 20
3.2 Adjustment for the coverage error at the population and household levels 22

4. Response errors 23

5. Processing errors 24

5.1 Quality indicators coming from imputation 26

Bibliography 32

Appendix A 33

Algebraic Notation

Appendix B 35

2001 Food Expenditure Survey Questionnaire (FE2)

Appendix C 42

2001 Food Expenditure Survey Diary (FE3) — Example for one day

Highlights

Sampling errors

- The coefficients of variation (CV) of the estimate of total average weekly expenditure per household vary between 1.9% and 3.1% at the regional level (five Canadian regions) and the CV is 1.1% at the national level.

- The coefficients of variation for average weekly expenditures per household by main summary categories of expenditures are generally less than 4% nationally and less than 9% regionally.

Non-response

- The non-response rate is 28.5%. It is due to refusals (17.0%) and households that could not be contacted (11.4%).

- The final non-response rate tends to increase with the level of urbanization. We observed a non-response rate of 19.2% at the rural level, and 30.0% in urban centres of one million inhabitants or more.

- Analysis of final response rates by strata of high-income and low-income geographic areas drawn from the sample design reveals that the non-response rate in high-income strata (37.2%) is higher than the rate in low-income strata (30.4%) and regular strata (27.8%).

Coverage errors

- Undercoverage of households is 7.9% at the national level and ranges between 5.0% and 10.4% at the regional level.

- Undercoverage of individuals is 6.9% at the national level, and ranges between 4.6% and 9.2% in each region.

- National slippage rates for children (0 to 6 years and 7 to 17 years) are quite different from the rates for other age groups. As a result, there is overcoverage of children and undercoverage of adults.

Response errors

- Data collection involves a personal interview with the completion of a paper questionnaire, and two weekly purchasing diaries that the respondent must fill in daily for two consecutive weeks. The estimate of expenditures obtained by the questionnaire is 8.3% higher than that obtained from the diaries.

Processing errors related to imputation

i) From the questionnaire (FE2)

- Only two sections are imputed: *Spending habits* (section B) and *Food and beverages while away from home overnight or longer during the previous month* (section C). For section B, 12.1% of the questionnaires are imputed. Most of them have only one or two variables out of five needing imputation. For section C, 4.2% of the questionnaires are imputed.

ii) From the diary (FE3)

- For each day, only the section *Food and beverages purchased from stores* is imputed. Of the 279,709 food expenditures reported, 2.2% required imputation. Most of the time, the imputation specifies the type of expenditure. For example, the respondent indicates Milk without specifying the type of milk (Fluid whole, low-fat, skim, etc.).

Introduction

The Food Expenditure Survey (FES) is a periodic survey conducted every four to six years. The survey collects data from Canadian households on food spending habits. Data are collected by a personal interview using a paper questionnaire (FE2), and by two weekly diaries of purchases (FE3) that the respondent must fill in daily during two consecutive weeks. The 2001 Food Expenditure Survey was conducted every month during the 2001 calendar year. The 2001 FES sample consisted of 8,414 households distributed throughout the ten provinces and in the cities of Whitehorse, Yellowknife and Iqaluit.

This survey is used to complete the Survey of Household Spending (SHS)[1] by gathering detailed data on food expenditures that cannot be collected using the SHS methodology. The Food Expenditure Survey and Survey of Household Spending are used to update the weightings used in the Consumer Price Index.

The FES, like all surveys, is subject to error despite all the precautions taken at the various stages of the survey to control them. Although there is no exhaustive measure of a survey's data quality, certain quality measures taken at various stages of the survey can provide the user with relevant information to ensure sound data interpretation.

This paper presents the quality indicators produced for the 2001 Food Expenditure Survey. It includes the usual quality indicators that are generally useful to users in interpreting the data, such as coefficients of variation, non-response rates, slippage rates and imputation rates.

The quality indicators have been classified according to the main types of error that can be found in a survey. Section 1 discusses the sampling errors, that is, errors due to the fact that the inferences made regarding the population as a whole are based on the data collected from a sample of the population and not from the whole population. The subsequent sections cover errors due to factors other than sampling. Non-response and coverage errors are discussed first in Sections 2 and 3. Response errors and processing errors are dealt with in sections 4 and 5 respectively.

This paper focuses on data quality. For a detailed description of the survey's methodology, consult Reference [1].

1. The SHS is an annual survey in which a sample of Canadian households is asked to report on all expenditures made during a calendar year. The results of the SHS are published in *Spending Patterns in Canada*, Catalogue No. 62-202.

1. Sampling errors

Sampling errors arise from the fact that the inferences made from the survey regarding the population as a whole are based on information that is gathered from a sample of the population and not from the whole population. In addition to the survey design and estimation method applied for the Food Expenditure Survey, the size of the sample and the variability of each characteristic are determining factors of sampling error. Characteristics that are rare or which are distributed with high variability in the population will have a larger sampling error than characteristics that are found more frequently or which are distributed with less variability in the population.

1.1 Measurement of sampling error

Standard error is a common measure of the sampling error. Standard error corresponds to the degree of variation in the estimate given that a specific sample has been chosen rather than some other from among all of the possible same-size samples under the same sample design. Since the FES uses a complex sample design and estimation method, the standard error is estimated using a resampling method known as the Jackknife technique. See Reference [2] for further details on this method.

The coefficient of variation (CV) is also a frequently used measure of estimate reliability. It simply states the standard error as a percentage of the estimate. Thus, if an estimate Y is obtained for a certain characteristic and the SE corresponds to the estimated standard error, then the CV will be (SE/Y) x 100.

Finally, the standard error or the coefficient of variation can be used to derive another measure of the accuracy of estimates – the confidence interval. This measure indicates the level of confidence there is that the real value of a certain observed characteristic in the population will be found within certain limits. An interval with a confidence level of 95% corresponds to an estimate obtained from the sample with ±2 standard errors: (Y ± 2 SE). This means that if the sampling were repeated many times, each sample would provide a different interval and 95% of the intervals would contain the real value of the characteristic. Similarly, if the sampling were repeated, the interval Y± SE would contain the real value in 68% of cases. Note that the confidence interval is also calculated from the CV in a similar manner, namely (Y ± 2 (CV x Y) / 100).

1.2 Coefficients of variation

Estimates of coefficients of variation are calculated for the estimates of several characteristics collected in the FES. The CVs for weekly estimates of average expenditure per household, reporting percentage (proportion in percentage of diaries having expenses for a particular expenditure) and average quantity per household are available at the national and regional level in the publication *Food expenditure in Canada* (see Reference [3]).

It should be noted that CV estimates do not take into account the fact that some data are imputed. Consequently, the CVs presented may underestimate the real

values. For most variables, the imputation rate is relatively low (see Section 5) and the CVs provided represent a good estimate of the real CVs. However, it is important to take into account both the CV and the imputation rate when examining the reliability of detailed food expenditures with a high imputation rate.

Table 1.1 gives an overview of the CVs of the estimates of average weekly expenditure per household at the regional level and at the national level for a few of the summary expenditure categories.

Table 1.1
Coefficients of variation (%) of average weekly expenditure per household for certain summary categories of expenditure, five Canadian regions and Canada (10 provinces)

Summary categories of expenditure	**Canada (10 prov)**	Atlantic	Quebec	Ontario	Prairies	British Columbia
Food purchased in stores	1.0	3.1	1.8	1.9	3.1	2.1
On trips overnight or longer	10.5	22.4	13.6	13.7	35.4	15.3
Locally and on day trips	1.0	2.9	1.8	1.8	2.3	2.1
Meat	1.5	4.5	2.6	2.6	3.9	3.5
Fish and other marine products	3.5	8.2	7.2	6.0	7.3	8.5
Dairy products and eggs	1.2	3.1	2.4	2.3	2.6	2.7
Bakery and other cereal products	1.2	3.0	2.0	2.3	2.5	2.6
Fruits and nuts	1.4	4.4	2.6	2.5	3.0	3.2
Vegetables	1.5	4.0	2.6	2.8	3.3	3.3
Condiments, spices and vinegar	1.9	5.5	3.6	3.6	4.3	4.5
Sugar and sugar preparations	2.3	5.6	5.0	4.2	4.5	5.4
Coffee and tea	2.9	7.1	5.4	5.0	8.3	6.4
Fats and oils	2.7	6.7	5.2	5.2	5.6	7.4
Other foods, materials and food preparations	1.7	5.1	3.3	3.2	4.2	3.7
Non-alcoholic beverages	2.0	5.2	3.8	3.6	4.9	5.3
Food purchased in restaurants	2.2	5.9	4.4	4.4	4.4	4.5
Total of weekly food expenditure	1.1	2.9	1.9	2.0	3.1	2.1

The CVs of the estimate of the total average weekly expenditure per household vary from 1.9% to 3.1% at the regional level and the CV is 1.1% nationally. The category of foods purchased in stores has CVs similar to those of the total, while the CVs of food purchased in restaurants are almost twice as high.

We see that the CVs for food purchased in stores on trips overnight or longer are much higher than the others. At the national level, the CV is 10.5% and at the regional level, they vary between 13.6% and 35.4%.

The quality of the estimates varies depending on the food category. The CVs for the "Fish and other marine products", "Coffee and tea" and "Fats and oils" categories are around 3% nationally and vary between 5% and 9% regionally. The nine other categories have CVs varying mainly between 1% and 2% at the national level and between 2% and 6% at the regional level.

1.3 Model for deriving an approximation of the CV

Estimates for various domains of interest (e.g. by size of area of residence) are available for the summary categories of expenditure in the publication *Food expenditure in Canada* (Reference [3]). Estimates for the various domains of interest for detailed categories of expenditure are also available by request from the Income Statistics Division. For operational reasons, it is not possible to produce the CVs for all the various levels of aggregation that might be of interest to users.

1.3.1 Approximation of the CV for domain estimates

However, it is possible to calculate an approximation of the CV using the relationship of the number of weekly diaries in which expenditures for an item were reported and the CV at an aggregate level. This relationship, based on the CV's tendency to grow proportionately to a decrease in the square root of the number of weekly diaries with a non-zero amount, is illustrated below.

Formula for approximating the CV for a domain (a sub-group of the population)

If CV(Y) represents the CV for the estimate of the average weekly expenditure per household of a certain characteristic for the whole population, then it is possible to calculate an approximation of the CV of the estimate of that characteristics for a domain (which can be considered a sub-group of the population, such as type of household, level of urbanization, etc.) using the following equation:

$$CV\ (Y_d) = CV\ (Y) \times \sqrt{\frac{n}{n_d}}$$

where

n = *the number of weekly diaries with a non-zero amount for the characteristic*

which can be obtained by calculating

$$\frac{\textit{the number of weekly diaries} \times \textit{the estimate of the percentage reporting}^2}{100}$$

n_d = *the number of weekly diaries with a non-zero amount for the characteristic in domain d*

which can be obtained by calculating

2. Proportion in percentage of diaries with a non-zero amount for the characteristic.

the number of weekly diaries in domain d x the estimate of the percentage reporting² in domain d

---

100

In general, the CV, the number of weekly diaries and the percentage reporting at the national level are used to calculate approximations for the various domains. In a case where we want to calculate the CV for a domain wholly contained in a single region (e.g. metropolitan area), it is preferable to use these values at the regional level since the regional CVs are published for the 2001 FES (Reference [3]). It is important to remember that the value of the CV obtained by this approach is only an approximation of the CV.

1.3.2 Approximation of the CV using microdata files

Users of microdata files can use another approach to derive an approximation of the CV of estimates, which will generally be more accurate than the one described in the previous section for CVs of detailed categories of expenditure. This approach is described in detail in the documentation accompanying the 2001 microdata files. It can only be used with the microdata files because both the data and the weights for each household are needed to calculate this approximation.

The paper on data quality of the 1997 Survey of Household Spending (Reference [4]) contains the results of the evaluation of the performance of these two methods of CV approximation.

1.4 Suppression of unreliable data in tables of estimates

Since the coefficient of variation is an indicator of data quality, we would like to use it to determine whether the estimates should be published. Estimates for which the CV is estimated at over 33% are not considered sufficiently reliable for publication. To facilitate tabulation, the suppression rule for food expenditure estimates is based on the number of weekly diaries with a non-zero amount for an expenditure category.

Based on the analysis of previous surveys, the CVs are usually below 33% when the number of weekly diaries with a non-zero amount for an item is greater than 40. Since this is an approximation rule, certain estimates may be published even if the CV is higher than 33% (type I error) and others will not be published even with a CV less than 33% (type II error). Table 1.2 below shows the value of these two errors for the CVs of estimates of average weekly expenditure per household for 365 food items.

Table 1.2
Type I and Type II error (%) using the 40 diaries rule

Domain	CV = 0	Type I error	Type II error	Number of CVs involved
Canada (10 provinces)	0.0	1.6	0.3	365
Region	0.5	0.6	7.7	1,825
Province	2.9	0.7	14.8	3,650
Metropolitan area	7.0	0.7	18.3	5,840
Level of urbanization	0.8	0.9	10.2	2,555
Household size	0.5	0.5	6.8	1,825
Household composition	3.2	0.9	12.3	2,920

The column "CV = 0" represents the proportion of estimates whose CVs are equal to zero. This occurs mainly because there is no weekly diary with a non-zero amount for the associated estimates.

The column "Type I error" represents the proportion of estimates that would be published under the 40 diaries rule, whose CVs are over 33%. This error is 1.6% at the Canada level and less 1% for the other studied domains[3].

The column "Type II error" represents the proportion of estimates, that would not be published under the 40 diaries rule, whose CVs are 33% or less.

2. Non-response

Errors due to non-response arise from the fact that certain potential respondents do not provide the necessary information or that this information is unusable. When the respondent has failed to respond to certain questions only, it is referred to as a partial non-response. In this case, the missing data are imputed. Errors associated with imputation are presented in Section 5 which deals with processing errors.

There is also another type of partial non-response: respondents who reported only one week of data. For the 10 provinces, there were 252 such cases.

In this section, non-response covers collection non-response, due mainly to an inability to contact the household or the refusal by the members of the household to participate in the survey.

3. This error is higher at the Canada level because of the proportion of published estimates compared to the other domains. This proportion is over 98% at the Canada level, while it is less than 85% for the other domains.

The main impact of non-response on data quality is that it can introduce a bias in the estimates if the characteristics of respondents and non-respondents differ and that difference impacts the characteristics examined. Non-response rates can be easily calculated but they give only an indication of data quality because they cannot be used to measure the size of the bias associated with the estimates. The level of non-response can be considered as an evaluation of the risks of bias in the estimates.

2.1 Non-response rates and vacancy rates

In the FES, because the units selected are dwellings, interviewers must first identify ineligible dwellings, that is, dwellings occupied by persons who are not part of the target population, dwellings that no longer exist (demolished, mobile home moved or dwelling converted into a business) and vacant dwellings (unoccupied, seasonal or under construction).

Among the eligible dwellings, the next step is to evaluate the proportion of households that did not respond to the survey, which is referred to as the collection non-response. This group includes households that refused to participate in the survey and households in which respondents could not be contacted because they were absent or because of special circumstances (language problem, illness or death).

Table 2.1 presents the non-response rates broken down by refusals and non-contacts. It also includes the vacancy rates. These rates are unweighted. They are provided at the national level and the regional level and for the three northern cities, by quarter and for the year.

Table 2.1
Non-response rates (%) and vacancy rates (%) by quarter and for the year, five Canadian regions, Canada (10 provinces) and three large cities in the Territories

Quarters		Canada (10 prov)	Atl.	Que.	Ont.	Pr.	B.C.	3 large territory cities	Whitehorse	Yellowknife	Iqaluit
January to December	Vacancy	7.5	12.4	7.4	6.2	8.4	5.8	13.9	16.8	10.7	14.2
	Non-response	28.5	20.4	22.8	33.2	23.6	37.0	32.4	41.1	18.2	39.8
	No contact	11.4	8.2	7.2	14.5	9.2	15.3	14.6	17.3	8.3	19.5
	Refusal	17.0	12.3	15.6	18.7	14.3	21.7	17.8	23.8	9.9	20.3
January to March	Vacancy	8.2	13.1	8.4	7.1	8.2	5.8	18.4	28.8	18.6	4.1
	Non-response	34.7	22.4	29.6	42.1	28.4	45.1	37.8	25.5	31.4	57.8
	No contact	15.0	9.3	10.4	20.4	11.4	19.8	17.5	12.8	7.8	33.3
	Refusal	19.7	13.1	19.2	21.7	17.0	25.3	20.3	12.8	23.5	24.4
April to June	Vacancy	7.7	9.2	7.7	6.0	10.2	7.1	12.8	9.8	13.3	16.3
	Non-response	29.7	20.1	23.6	32.9	30.0	36.8	39.7	55.8	26.3	29.0
	No contact	11.3	6.4	7.8	13.7	10.8	14.7	18.2	21.2	15.8	16.1
	Refusal	18.5	13.7	15.8	19.3	19.3	22.2	21.5	34.6	10.5	12.9
July to September	Vacancy	7.3	11.5	6.5	6.9	8.1	5.3	8.5	9.5	3.3	14.3
	Non-response	27.0	17.3	18.8	31.1	28.1	33.7	28.9	42.1	15.1	28.1
	No contact	10.5	6.1	5.3	12.3	12.4	13.8	13.4	19.3	9.4	9.4
	Refusal	16.5	11.2	13.5	18.9	15.7	19.9	15.5	22.8	5.7	18.8
October to December	Vacancy	6.9	15.3	7.0	4.8	7.2	5.0	15.2	18.3	6.5	23.8
	Non-response	21.3	21.4	18.7	25.9	6.1	31.8	23.1	39.1	2.0	36.0
	No contact	8.3	10.7	5.1	11.1	1.7	12.7	9.1	15.2	2.0	12.0
	Refusal	13.0	10.7	13.6	14.8	4.4	19.1	14.0	23.9	0.0	24.0

The non-response rate for the year in Canada (10 provinces) is 28.5%. This rate includes refusals (17.0%) and households could not be contacted (11.4%). For all regions except the North, regardless of quarter, refusals represent the main reason for non-response. The three northern cities combined have a non-response rate for the year of 32.4%. This rate includes refusals (17.8%) and households that could not be contacted (14.6%). In the case of Yellowknife and Iqaluit, for certain quarters, the main reason for non-response is non-contact.

Non-response rates tend to decrease from one quarter to the next. We note that these rates are especially high in the first quarter in Ontario and British Columbia. They reflect collection problems encountered in the first quarter.
The Ontario and British Columbia regions have the highest non-response rates. For the year as a whole, their non-response rates are respectively 33.2% and 37.0%, while the three other regions have non-response rates between 20.4% and 23.6%.

Among the three northern cities, Yellowknife had the lowest non-response rate at 18.2% for the year. Whitehorse and Iqaluit have non-response rates for the year of 41.1% and 39.8% respectively.

The vacancy rates are shown in Table 2.1, but it should be remembered that vacant dwellings do not contribute to the sample bias to the extent that they are correctly identified. The analysis of the vacancy rates can reveal dwelling identification problems associated with collection. The vacancy rate for the 2001 FES is 7.5% for Canada (10 provinces) and 13.9% for all of the three Northern cities.

2.2 Non-response according to the level of urbanization

Non-response varies according to the level of urbanization. The various rates for Canada (10 provinces) are given by level of urbanization, by quarter and for the year in Table 2.2.

Table 2.2
Non-response rates (%) and vacancy rates (%) by level of urbanization, by quarter and for the year, Canada (10 provinces)

Category of urbanization		TOTAL	URBAN	1,000,000 or more	500,000 to 999,999	250,000 to 499,999	100,000 to 249,999	30,000 to 99,999	Less than 30,000	RURAL
January to December	Vacancy	7.5	4.4	3.4	3.4	5.1	5.6	5.0	6.6	22.6
	Non-response	28.5	30.0	36.4	28.5	28.9	22.2	27.9	20.9	19.2
	No contact	11.4	12.2	16.1	11.3	10.8	6.7	10.1	8.0	7.0
	Refusal	17.0	17.8	20.3	17.2	18.0	15.4	17.8	12.8	12.2
January to March	Vacancy	8.2	4.8	4.0	3.6	8.0	5.2	5.4	5.2	23.8
	Non-response	34.7	36.5	45.0	38.2	35.3	24.6	30.5	26.0	24.3
	No contact	15.0	15.8	22.0	17.5	16.3	6.1	9.3	9.2	9.8
	Refusal	19.7	20.6	23.0	20.7	18.9	18.4	21.2	16.8	14.5
April to June	Vacancy	7.7	4.8	4.3	4.1	4.6	8.4	2.1	6.2	21.1
	Non-response	29.7	31.1	36.3	31.3	23.9	28.5	31.6	21.3	22.0
	No contact	11.3	12.1	16.3	10.8	6.1	8.2	11.8	7.9	6.6
	Refusal	18.5	19.0	19.9	20.5	17.8	20.3	19.9	13.4	15.4
July to September	Vacancy	7.3	4.6	2.9	3.1	5.1	5.3	6.8	10.0	20.5
	Non-response	27.0	28.6	31.8	29.0	29.9	19.3	31.4	23.2	16.9
	No contact	10.5	11.1	13.1	11.2	9.1	6.8	14.0	8.2	6.5
	Refusal	16.5	17.5	18.8	17.8	20.7	12.5	17.4	14.9	10.4
October to December	Vacancy	6.9	3.3	2.4	2.7	2.2	4.1	5.9	5.5	24.5
	Non-response	21.3	22.8	32.0	13.0	25.6	16.4	17.5	11.5	11.9
	No contact	8.3	9.0	12.6	4.0	10.8	6.2	5.6	6.5	4.2
	Refusal	13.0	13.8	19.4	9.1	14.8	10.2	11.9	5.0	7.7

The non-response rate for the year tends to increase with the level of urbanization. There is a difference of 17.2% between the categories of 1,000,000 or more (36.4%) and Rural (19.2%). Refusals are the main cause for non-response, except in one case: the less than 30,000 group in the last quarter.

An examination of vacancy rates by level of urbanization reveals that the vacancy rate is higher in the rural regions (22.6%) than in low population urban areas (6.6%). This same phenomenon is also found in the Labour Force Survey (LFS) and is no doubt due to the large number of seasonal dwellings in rural settings. Since the FES sample is more concentrated in high population urban areas than the LFS, we can expect to see a national vacancy rate slightly lower than that of the LFS. Effectively, the 2001 FES has a vacancy rate of 7.5% while the rate for the LFS for the same year is 12%.. For further details on the LFS methodology, see Reference [5].

2.3 Non-response by income strata

It is impossible to compare non-response rates by income level because this information is not given by the non-respondents. However, the LFS sample design used for the FES was designed in part to establish, in nine major cities, strata of geographical areas where the average household income exceeds $100,000 and in seven major cities, strata composed of apartments inhabited by households with an average income below $20,000. This is a small number of strata and it accounts for only a small number of dwellings in the FES sample (about 530 and 75 for high-income and low-income strata respectively, or slightly more than 6% of the sample).

Table 2.3
Comparison of non-response rates (%) and vacancy rates (%) of high income and low income strata in relation to other strata, by quarter and for the year, Canada (10 provinces)

Type of stratum by income		TOTAL	High income	Regular	Low income
January to December	Vacancy	7.5	3.2	7.8	6.7
	Non-response	28.5	37.2	27.8	30.4
	No contact	11.4	16.0	11.1	13.0
	Refusal	17.0	21.2	16.7	17.4
January to March	Vacancy	8.2	5.6	8.4	0.0
	Non-response	34.7	44.3	34.0	28.6
	No contact	15.0	24.2	14.4	0.0
	Refusal	19.7	20.1	19.6	28.6
April to June	Vacancy	7.7	0.8	8.2	5.3
	Non-response	29.7	41.0	28.8	38.9
	No contact	11.3	18.0	10.8	5.6
	Refusal	18.5	23.0	18.0	33.3
July to September	Vacancy	7.3	3.3	7.4	15.8
	Non-response	27.0	35.7	26.4	25.0
	No contact	10.5	13.4	10.2	25.0
	Refusal	16.5	22.3	16.3	0.0
October to December	Vacancy	6.9	2.5	7.2	4.3
	Non-response	21.3	25.6	20.9	28.6
	No contact	8.3	6.0	8.4	19.0
	Refusal	13.0	19.7	12.6	9.5

Looking at the year, the highest non-response rate is in the high-income strata (37.2%), followed by the low-income strata (30.4%) and regular strata (27.8%). This same order occurs for non-contacts and for refusals. The main cause of non-response is refusals.

Still looking at the year, the highest vacancy rate is found among the regular strata (7.8%).

In the quarters, non-response rates vary from quarter to quarter. In particular, refusal rates in the first two quarters are highest in the low-income strata.

2.4 Adjustment for non-response

To offset non-response, the FES weights are increased by the inverse of the weighted response rate within certain defined groups in each region by quarter. The defined groups in each region include mainly the high-income strata, certain metropolitan areas and the various levels of urbanization. The weighted rates differ from the rates presented in this section because they take into account the design weight of each household. An algebraic description of the adjustment for non-response is given in Appendix A.

The adjustment of the weights for non-response makes it possible to take into account differences in the level of non-response by defined group, and will also reduce the bias to the extent that the characteristics of the respondents and non-respondents are similar for a given group.

3. Coverage errors

The population in question, referred to as the target population, was defined during the survey's design. It is important to reiterate the definition of this target population for the FES because a good understanding of the target population is necessary for a sound interpretation of the survey data. It is important to point out that the FES uses the survey frame of the Labour Force Survey (LFS).

Target population

> The target population corresponds to individuals living in private households in Canada in the ten provinces and the cities of Whitehorse, Yellowknife and Iqaluit. Persons living full-time in institutions, such as prisons, chronic care hospitals and residences for senior citizens, and persons who are members of religious and other communal colonies, members of the Armed Forces living in military camps, and individuals living full-time in hotels or rooming houses are excluded. Also excluded are official representatives of foreign countries living in Canada and their families, and persons living on Indian reserves and crown lands.

We did not collect data from persons temporarily living away from their families (for example, students at university) because the data would be gathered from their families if selected in the sample.

Furthermore, in 2001, for practical reasons mainly related to the fact that the size of the sample was much smaller than in previous surveys, households residing in dwellings located in remote areas were excluded from the survey. These regions, which are sparsely populated for the most part, are located mainly in the northern parts of certain provinces. They are difficult to contact and data collection is very costly for these households.

The survey therefore covers about 97% of the population in the 10 provinces. In the Yukon, Northwest Territories and Nunavut, the survey covers only the cities of Whitehorse, Yellowknife and Iqaluit, which represent 62%, 44% and 17% of the population of their respective territories based on the 1996 Census.

Coverage errors are the result of inadequate representation of the target population from the units of the survey frame. Certain units of the target population may be omitted from the survey frame, which creates a situation of undercoverage. Other units that are not in the target population may be included by mistake or certain units may be included more than once. These units are responsible for overcoverage.

3.1 Undercoverage and overcoverage: slippage rates

In the FES, the selection of the sample is done using a list of dwellings in each selected cluster. The omission of dwellings during the creation of the list, new dwellings that are added between the creation of the list and the visit of the interviewers (mainly in developing areas), and the incorrect classification of vacant dwellings contribute to undercoverage. Including dwellings that would not be within the parameters of the cluster is a source of overcoverage. Similarly, errors can slip in during the gathering of the data, when identifying persons who are members of the selected household. These errors also contribute to undercoverage and overcoverage.

A good representation of the target population is essential to producing realistic estimates of expenditures. The number of people per household is also an important characteristic in the estimate of average household expenditure. It is therefore necessary for the sample not only to properly represent the individuals in the target population, but also the distribution of households by size.

In general, there is net undercoverage of the number of people and the number of households in the FES that is corrected by an adjustment of the weights using auxiliary data, which are based on post-censal demographic estimates. The slippage rate (see algebraic description in Appendix A) is a measure of the percentage of difference between the estimates from these auxiliary data and the estimates from the survey calculated using weights not adjusted with these data.[4] Slippage rates by age group at the national and regional level are presented in

4. Sub-weights are used, which are design weights adjusted for non-response (see Appendix A).

Table 3.1, while slippage rates by size of household are found in Table 3.2. A positive rate corresponds to overcoverage of the number of people or households in the survey.

Table 3.1
Slippage rates (%) by age group, five Canadian regions and Canada (10 provinces)

Age	Canada (10 provinces)	Atlantic	Quebec	Ontario	Prairies	British Columbia
0-6 yrs	4.7	5.5	8.6	3.0	6.1	0.9
7-17 yrs	1.5	2.2	-3.9	2.7	0.9	7.4
18-34 yrs	-13.9	-19.8	-10.7	-17.9	-5.7	-15.3
35-54 yrs	-7.7	-7.7	-11.0	-7.2	-4.9	-6.4
55-64 yrs	-6.5	-19.7	1.0	-7.3	-6.4	-11.1
65 yrs and +	-8.9	-6.6	-6.4	-5.8	-16.1	-15.0
Total	-6.9	-9.2	-6.6	-7.3	-4.6	-7.6

For the 2001 FES, the rate of undercoverage of the population is 6.9%. An analysis of Table 3.1 at the age group level reveals that, both nationally and regionally, the slippage rates for children (0 to 6 years and 7 to 17 years) are very different from those of the other age groups. In fact, among children, we find overcoverage or slight undercoverage, while among adults, there is always undercoverage, except in Quebec for the 55 to 64 age group.

The highest undercoverage rate at the national level is found in the 18 to 34 age group (13.9%). At the regional level, excluding the Prairies, this group has a high undercoverage rate. We also find a high undercoverage rate in the 35 to 54 age group in Quebec (11.0%), in the 55 to 64 age group for the Atlantic region (19.7%), and in the 65 and older age group in the Prairies and British Columbia (16.1% and 15.0% respectively).

Table 3.2
Slippage rates (%) by size of household, five Canadian regions and Canada (10 provinces)

Geography	Households	One-person households	Two-person households	Three or more person households
Canada (10 provinces)	-7.9	-16.9	-3.3	-5.8
Atlantic	-8.2	-13.8	2.9	-14.0
Quebec	-7.4	-17.2	0.4	-6.1
Ontario	-8.5	-16.5	-5.1	-6.5
Prairies	-5.0	-14.1	-2.4	-1.6
British Columbia	-10.4	-22.2	-9.9	-3.1

At the national level, undercoverage of the number of households was 7.9%. This slippage rate is 1 percentage point higher than the slippage rate for the population which is 6.9%, see table 3.1.

At the regional level, the undercoverage rate varies between 5.0% and 10.4% for the number of households. With the exception of the Atlantic Region, the undercoverage rate is much higher among households with one person than among households with two or three or more persons. In the case of the Atlantic region, households with one person and those with three or more persons have similar undercoverage rates, specifically 13.8% and 14.0% respectively.

3.2 Adjustment for the coverage error at the population and household levels

To correct the problem of the sample's representativeness shown in Table 3.1 and to reduce the resulting bias, the survey data is adjusted during weighting using demographic estimates for the age groups defined in this table, for each Canadian region. This adjustment considerably reduces the bias caused by coverage errors but it does not eliminate the bias entirely if the characteristics of the individuals omitted or non-respondents differ from those respondents included for the same age group in a given region.

It should also be noted that the effectiveness of the coverage adjustment using demographic estimates depends largely on the quality of those demographic estimates and their accuracy in representing the survey's target population. The demographic estimates are not free from error. They are post-censal estimates based on the 1996 census population counts, adjusted for net undercoverage and which take into account recent statistics on migration, births, mortality, etc. These demographic estimates are adjusted to reflect certain exclusions specific to household surveys, such as persons living in institutions. Conceptually, they differ slightly from the FES target population by including persons living in collective households that are not institutions, such as members of communal colonies and individuals living full-time in hotels or rooming houses. However, this difference is considered negligible because these individuals represent less than 0.4% of the Canadian population.

To correct the problem of the sample's representativeness in terms of the number of households by size, shown in Table 3.2, the survey's data are adjusted using auxiliary data. To offset the bias caused by inadequate representation of households, the FES weights are adjusted to reflect post-censal estimates of the number of households by size for each Canadian region. However, this does not necessarily eliminate the bias if the characteristics of the households not interviewed (omitted or non-respondent) differ from those of the respondent households of the same size. As with the demographic estimates of population, the effectiveness of the adjustment depends on the quality of the auxiliary data on the number of households.

In addition to the demographic estimates mentioned earlier, two other sets of auxiliary data are used to adjust the survey data during weighting in order to improve representativeness. The first set of data is used to control the number of

children and adults in certain large cities. The second is designed to control the number of single-parent households and couples with children households by Canadian region.

For further information on the methodology of the adjustment, including that for northern cities, see Reference [1].

4. Response errors

Response errors correspond to inaccurate responses to questions. They can be attributed to a variety of factors including a questionnaire that needs improving, incorrect interpretation of questions by interviewers or respondents, and errors in respondents reported data.

Respondents are required to record their expenditures on an ongoing basis in the two weekly diaries given to them. However, errors can occur when respondents make transcription errors, do not use the right terms for products or guess the amounts when there is no price sticker or quantity of a product.

The survey has a number of features to help respondents give information that is as accurate as possible. During personal visits, the interviewer gives indications of how to properly complete the diary. Sometimes, he helps the respondent record missing data. In addition, for 2001, the respondents were strongly encouraged to attach their receipts to the page in the diary corresponding to the date of the purchase. There is another new element in 2001, namely, the notebook. All members of the household aged 10 years and older were given a small notebook to carry around at all times in which to record products at the time of purchase.

It is also felt that the burden imposed on the respondent, which consists mainly of reporting a wide variety of items over a two-week period, may lead to respondent fatigue and may affect the quality of the responses obtained. Due to this burden related to the diary (FE3), we make the assumption that the estimate of food purchased in stores according to the questionnaire is more reliable than the corresponding estimate from the diary.

In 2001, the estimate of expenditures obtained from the questionnaire (FE2) is 8.3% higher than that obtained from the diary data (FE3). Based on the assumption mentioned above, the values for food purchased in stores locally and on one-day trips in each household were multiplied by an adjustment factor of 1.083.

Although response errors are considered an important source of error, they remain the aspect of data quality that is the most difficult to measure. It is generally necessary to undertake costly special studies to try to measure them. We try to reduce them by implementing different procedures during the processing such as editing for example.

5. Processing errors

Errors can arise at all steps of data processing The main stages of data processing are coding, data entry, editing, imputation of partial non-response and weighting. In the FES different procedures are applied at each stage in order to minimize processing errors and the survey estimates are compared with other data sources prior to release. Errors related to the adjustments made at the weighting stage have been described in sections 2 and 3. The other types of processing errors are covered in this section.

Coding is an important step in the FES. All the described items bought must be coded. This is done at the head office of Statistics Canada by coders using a software developed in the Income Statistics Division to help them to choose the appropriate food codes. During coding, each questionnaire and diary goes through a series of manual edits and imputation. Some automated edits are also present during the coding, such as those identifying inconsistencies between codes and units of measure. Also, some questionnaires selected randomly are verified by the senior coders.

In terms of data entry, the workload for a keyer is defined as 20 questionnaires split in 4 batches of 5. A batch is selected randomly. If the data quality of the batch is acceptable, then the workload is accepted, if not then the 3 other batches are checked.

As mentioned previously, during coding the questionnaires and diaries are subjected to manual and automated edits and manual imputation. An automated edit process is carried out after each questionnaire and diary passes these manual steps.

This automated edit process still identified some data that need to be imputed, most of which was missing data. An example of missing data on the questionnaire (FE2) is having an expenditure for restaurants while away from home overnight or longer during the previous month but the number of meals purchased is missing. An example of missing data on the diary (FE3) is having raspberries as a purchased item, but the type of packaging (fresh, frozen, etc.) is missing. The imputation at this stage is automated. More than one method of imputation is used, but the basic one in 2001 is the nearest neighbour method.

This technique involves forming groups of similar households based on certain criteria (e.g., province of residence, quarter of data collection). Within those groups, each household requiring imputation (recipient) is matched to a household that has a complete questionnaire (donor) and resembles the other most closely with respect to certain characteristics (e.g. income, household size, etc.). The donor's data are imputed to the recipient as long as they satisfy the edit requirement for consistency with the data reported by the recipient. For example, if the recipient has indicated the buying of prepared food then the imputed expenditures for this variable must be greater than zero.

If a donor cannot be found for some recipients during this first step, then the donor imputation method is repeated without including the household income

criteria, which is considered less significant than the other criterion. Many donor imputation steps are required to be able to impute all recipients. For a few cases, the most significant condition, which is the quarter, is also dropped to find a donor for the left over cases. For more details on the different donor imputation steps, see Reference [6].

In the case of the questionnaire (FE2), two sections are imputed: *Spending habits* (section B) and *Food and beverages while away from home overnight or longer during the previous month* (section C). These sections are imputed independently since they are considered fairly independent from each other. The nearest neighbour approach is used for section B, but for section C, it is sometimes necessary to use other methods due to the difficulty in finding an appropriate donor. For example, some respondents have reported expenditures in restaurants but not the number of meals purchased or vice versa. These cases are imputed using the mean imputation method. The imputed mean used is the one at the national level.

In the case of the diary (FE3), for each day, only the section *Food and beverages purchased from stores* is imputed. To impute this section, independent imputation runs are carried out based on the grouping of missing information needing imputation. Generally the imputation is done at the entry level but in some cases it is done at the day level. These cases occur when the respondent has only listed the total expenditures on food purchased from stores on a given day.

To impute at the entry level, the respondent must have provided additional information. For example, the respondent has purchased beef without specifying which kind of beef: ground, stewing, etc. In this case, the unit price is considered the most important matching criteria in finding a donor. If the unit price is missing then total expenditure is used along with other variables for matching.

Some reported entries are broad such as *Vegetables* for example. For the first time, in 2001, these entries are coded and imputed using the nearest neighbour method. Sometimes, a donor cannot be found for some recipients because of a high expenditure reported with the entry. To impute them, we sum for each potential donor all the expenditures in a day with the appropriate food entries (all vegetables for example). When a donor is found with the sum of expenditures matching the recipient expenditures, then all its appropriate food entries are assigned to the recipient.

When the respondent has only listed the total expenditures on food purchased from stores in any day, the imputation is done at the day level. The recipient day's expenditure is matched to the donor total day's expenditure. Only those days are considered as donors if they do not have any imputation at the entry level. All entries of a selected donor's day are assigned to the recipient record. The recipient total expenditures will be overwritten with the donor total expenditures when the two numbers do not match exactly.

Finally, note that households providing only one week of data are not imputed for the week they did not report.

The bias caused by imputation of partial non-response is difficult to evaluate. It depends on the differences between respondents and non-respondents as well as the ability of the imputation method to produce unbiased estimates. However, the percentages of households having imputed data and the imputation rates indicate the importance of partial non-response. The impact of imputation is a good indicator of the potential bias of estimates. These quality indicators are presented in the following section.

5.1 Quality indicators coming from imputation

A preliminary indication of the magnitude of partial non-response is the proportion of households requiring imputation and the number of variables imputed by household. Other preliminary indications are the imputation rates and the impact of imputation. This impact is defined as a measure of the impact of the imputed values on the estimates. An algebraic description of the impact of imputation is given in Appendix A.

The households' responses can be divided into two categories: those collected from the questionnaire (FE2) and those collected from the diaries (FE3). The different quality indicators for the questionnaire (FE2) are presented in the next sub-section and for the diaries (FE3) in the following sub-section.

5.1.1 Questionnaire (FE2)

Only two sections are imputed on the FE2: *Spending habits* (section B) and *Food and beverages while away from home overnight or longer during the previous month* (section C). There are up to five variables that can be imputed in section B (questions 1 to 4 and 4.1) and up to 50 in section C (questions 1 to 6).

The percentage of respondent households[5] that required imputation for section B is presented in Table 5.1 at the regional and national levels. The table is broken down by the number of imputed variables (out of 5) for a household.

Table 5.1
Percentage of imputed households for Section B in the questionnaire (FE2), five Canadian regions and Canada (10 provinces)

Geography	Number of variables imputed			Total
	1	2	3 to 5	
Canada (10 provinces)	0.8	10.8	0.6	12.1
Atlantic	0.4	8.4	0.0	8.8
Quebec	0.5	3.6	0.0	4.1
Ontario	0.6	15.4	0.7	16.6
Prairies	1.2	13.0	1.2	15.5
British Columbia	1.1	11.4	0.8	13.3

5. Respondent households correspond to all households living in eligible dwellings, excluding households who could not be contacted, or who refused to participate in the survey.

Table 5.1 indicates that nearly 88% of the households at the Canada level do not need imputation in section B. Most of the households requiring imputation have two variables to be imputed, usually questions 4 and 4.1 which ask for information about prepared food expenditures for occasions. The regions of Ontario (16.6%), Prairies(15.5%) and British Columbia (13.3%) have the highest percentages while in Quebec the percentage is the lowest (4.1%).

The impacts of imputation calculated for the 4 quantitative variables in section B are presented in Table 5.2 at the regional and national levels. This impact of imputation is defined as the total weighted imputed values divided by the total weighted values (including the imputed and non-imputed values).

Table 5.2
Impacts of imputation (%) for Section B in the questionnaire (FE2), five Canadian regions and Canada (10 provinces)

Geography	Food and other groceries (Question 1)	Non-food (Question 2)	Bulk (Question 3)	Prepared food for occasions (Question 4.1)
Canada (10 provinces)	0.8	1.0	0.0	12.2
Atlantic	0.4	0.7	0.0	23.9
Quebec	0.2	0.3	0.0	1.2
Ontario	0.7	0.7	0.0	15.5
Prairies	1.6	2.5	0.0	14.5
British Columbia	1.8	0.9	0.0	4.5

Table 5.2 indicates that the variable "Prepared food expenditures for occasions" (Question 4.1) is the one having the highest impact. It is the variable in this section having the highest imputation rate and also very few households have reported such expenditures. At the Canada level, this variable has an impact of imputation of 12.2% while the three other variables have an impact of imputation of less than 1.0%. At the regional level, if we examine only the impact of imputation for "Prepared food expenditures for occasions", the Atlantic region (23.9%) has the highest one followed by Ontario (15.5%) and the Prairies (14.5%). Quebec has the lowest impact of imputation with 1.2%.

The percentage of respondent households that required imputation for section C is presented in Table 5.3 at the regional and national level. The table is broken down by the number of imputed variables (out of 50) for a household.

Table 5.3
Percentage of imputed households for Section C in the questionnaire (FE2), five Canadian regions and Canada (10 provinces)

Geography	Number of variables imputed					Total
	1 to 10	11 to 20	21 to 30	31 to 40	41 to 50	
Canada (10 provinces)	1.6	0.0	2.1	0.0	0.4	4.2
Atlantic	1.3	0.0	1.8	0.0	0.0	3.1
Quebec	0.9	0.0	1.3	0.0	0.2	2.3
Ontario	0.9	0.1	2.1	0.0	0.8	3.9
Prairies	2.9	0.0	2.6	0.0	0.4	5.8
British Columbia	2.0	0.0	3.1	0.1	0.3	5.6

Table 5.3 indicates that at the Canada level, more than 95% of households do not require imputation in section C. Half of the households needing imputation have 21 to 30 variables to be imputed (mainly question 4 asking the food and non-alcoholic beverages purchased from restaurant while away from home overnight or longer involving 28 items to report) and most of the other half has 1 to 10 variables to be imputed. At the regional level, nearly 6% of the households in the Prairies and in British Columbia need imputation. Quebec has the lowest percentage, which is 2.3%.

The impacts of imputation calculated for the 31 quantitative variables excluding question 1 and 2[6] in section C is presented in Table 5.4 at the regional and national levels. Since there is a high number of variables instead of presenting the impact of imputation for each of them, the number of variables by category of impact of imputation (%) is presented.

Table 5.4
Impacts of imputation (%) for Section C in the questionnaire (FE2), five Canadian regions and Canada (10 provinces)

Geography	Number of variables by range of imputation impact (%) for Section C in the FE2					
	0 to 5%	5 to 10%	10 to 15%	15 to 25%	25 to 35%	35% or over
Canada (10 provinces)	8	15	5	2	1	0
Atlantic	13	8	5	2	1	2
Quebec	22	4	2	3	0	0
Ontario	15	4	7	1	3	1
Prairies	12	8	3	3	2	3
British Columbia	12	3	6	8	2	0

Table 5.4 indicates at the Canada level that 23 out of 31 variables have an impact of imputation of less than 10%. There are only 3 variables that have

6. The impact of imputation was calculated only for variables representing food expenditures or numbers of meals.

impacts over 15% and these all fall under the category "Other" in question 4: these are for *Lunches* (33.3%), *Dinners* (22.5%) and *Breakfasts* (21.7%). At the regional level, British Columbia is the region having the highest number of variables (10) with impacts of imputation of over 15% followed by the Prairies with 8 variables. Quebec has the lowest impacts.

5.1.2 Diary (FE3)

For each day, only the section *Food and beverages purchased from stores* is imputed on the FE3.

Of the 5,999 responding households, 279,709 expenditures are reported and 2.2% of them required imputation. Most of the time, the imputation specifies the type of expenditure. For example, if the respondent indicates Milk, without specifying the type of milk, this entry will be imputed through either Fluid whole milk, Low-fat milk (1%), Low-fat milk (2%), Fluid skim milk or Specialty milk products.

The percentage of respondent households that required imputation for the type of expenditure in the diaries (FE3) is presented in Table 5.5 at the regional and national levels. These proportions are high (38% at the national level) because each household has to list a large number of entries during the two-week reporting period. Any household having an imputation in any of these entries is considered to be an imputed household.

Table 5.5
Percentage of imputed households in the diaries (FE3), five Canadian regions and Canada (10 provinces)

Geography	Number of imputations required			Total
	1	2	3 or more	
Canada (10 provinces)	13.1	8.3	16.6	38.0
Atlantic	14.0	6.0	9.5	29.5
Quebec	16.1	10.8	18.8	45.6
Ontario	12.3	9.2	21.9	43.4
Prairies	11.6	6.5	12.3	30.5
British Columbia	11.9	7.4	14.8	34.2

Table 5.5 indicates at the national level that 38.0% of households have some imputed data in their diaries. It is worth noting, however, that a third of these households involve only one imputed entry while the average number of entries per household is 49. This average is based only on those households that reported expenses in diaries. At the regional level Quebec (45.6%) and Ontario (43.4%) have the highest percentages, but for households with 3 imputations or more, Ontario has a percentage of imputed households higher than Quebec. Atlantic has the lowest household percentage with 29.5%.

The type of food for each of the 279,709 reported expenditures is coded. This coding for the FES 2001 produced 205 different codes. The imputation rates and

the impacts of imputation for the 11 most reported expenditures are presented in Table 5.6 and Table 5.7 respectively. These imputation rates and impacts of imputation are calculated by type of expenditure. Both tables are at the regional and national levels. The tables are ordered by having the most frequently reported expenditures listed first.

Table 5.6
Imputation rates (%) in the diaries (FE3), five Canadian regions and Canada (10 provinces) for the 11 most frequently reported expenditures

Food description	Canada (10 prov)	Atlantic	Quebec	Ontario	Prairies	British Columbia
Bread	3.0	0.5	2.1	4.2	4.3	2.4
Carbonated beverages	1.6	0.1	1.4	2.1	3.2	0.1
Unsweetened rolls and buns	2.8	1.0	2.1	3.4	4.3	1.4
Bananas and plantains - fresh	3.7	0.7	3.5	5.2	4.4	2.4
Low-fat milk (2%)	15.6	15.6	13.8	19.9	14.7	11.7
Chicken (including fowl) - fresh or frozen	5.3	0.7	4.2	8.5	4.4	4.6
Cookies and sweet biscuits	4.0	0.6	2.2	7.5	4.9	2.1
Breakfast cereal (except infant cereal)	2.1	0.4	1.2	2.4	4.4	1.3
Soup - chilled, frozen or canned	2.2	0.3	2.2	2.5	3.6	1.0
Tomatoes - fresh	6.5	0.0	5.6	11.3	5.8	2.9
Eggs	2.1	0.4	2.2	2.2	3.5	1.4

At the Canada level, table 5.6 indicates that most of these expenditures have an imputation rate lower than 5%. The highest imputation rate is for 2% Low-fat milk with 15.6%. The imputation of this type of expenditure occurs most of the time with respondents that have reported an expenditure for milk but not specified the type of milk.

At the regional level, the highest imputation rates for these various expenditures are in Ontario and the Prairies. The Atlantic region is the one having the lowest rates, which are 1% or less, if we exclude the 2% Low-fat milk expenditure.

In Table 5.6, Quebec is one of the regions having low imputation rates for the 11 most reported expenditures even though it had the highest percentage of imputed households requiring at least one imputation rate as shown in Table 5.5. This can be explained by the distribution of the respondents' missing data. Quebec has more respondents missing at least one piece of information in their diaries, but this missing information is spread across the various types of expenditures to a larger degree in Quebec than in the other regions.

Table 5.7
Impacts of imputation (%) on the expenditure variable in the diaries (FE3), five Canadian regions and Canada (10 provinces) for the 11 most frequently reported expenditures

Food description	Canada (10 prov)	Atlantic	Quebec	Ontario	Prairies	British Columbia
Bread	3.6	0.4	2.0	5.7	4.3	2.8
Carbonated beverages	2.0	0.1	1.3	3.2	2.4	0.1
Unsweetened rolls and buns	3.9	1.5	1.8	5.8	5.0	1.2
Bananas and plantains - fresh	4.6	1.1	4.0	6.0	4.9	3.8
Low-fat milk (2%)	16.9	14.0	13.6	21.1	16.7	13.2
Chicken (including fowl) - fresh or frozen	7.5	0.3	2.9	13.4	4.2	4.2
Cookies and sweet biscuits	5.5	1.9	2.1	10.2	5.1	1.6
Breakfast cereal (except infant cereal)	2.1	0.6	1.1	2.7	3.9	1.1
Soup - chilled, frozen or canned	2.4	0.6	2.0	3.8	2.9	0.4
Tomatoes - fresh	11.6	0.0	7.0	20.6	5.5	3.1
Eggs	2.5	0.6	1.8	2.8	4.2	2.1

Table 5.7 indicates that at the Canada level, 7 of these expenditures have an impact of imputation based on the expenditure variable lower than 5%, 2 have impacts between 5% and 8% and 2 have impacts of 11% and 17%. At the regional level, the highest impacts of imputation are found mostly in Ontario. The impacts of imputation for the Atlantic region, Quebec and British Columbia are all below the national figures.

For more detailed information on the imputation, see Reference [6].

Bibliography

[1] Auger, S., Khan, K. and Tremblay J. (2004), *Methodology of the Food Expenditure Survey*, Catalogue no. 62F0026MIE-2005004, Household Survey Methods Division, Statistics Canada.

[2] Wolter, K.M. (1985), *Introduction to Variance Estimation*, Springer-Verlag New-York Inc.

[3] *Food Expenditure in Canada (2001)*, Catalogue no. 62-554-XIE.

[4] *1997 Survey of Household Spending – Data Quality Indicators*, Household Survey Methods Division, Internal Document, Statistics Canada.

[5] *Methodology of the Canadian Labour Force Survey*, Catalogue no. 71-526-XPB.

[6] Khan, K. (2002), *Imputation of the 2001 Food Expenditure Survey*, Household Survey Methods Division, Internal Document, Statistics Canada.

Appendix A

Algebraic Notation

1. Adjustment for non-response

The sub-weight (i.e. the design weight adjusted for non-response) of a household k, written w_k^{NR}, is

$$w_k^{NR} = \pi_k^{-1} * \frac{1}{rate_g} \quad with \quad rate_g = \frac{\sum_{k \in s_{g,r}} \pi_k^{-1}}{\sum_{k \in s_{g,r}} \pi_k^{-1} + \sum_{k \in s_{g,nr}} \pi_k^{-1}}$$

where

$s_{g,r}$ is the set of respondent households in non-response group g,

$s_{g,nr}$ is the set of non-respondent households (refusals, no contacts) in non-response group g, and

π_k^{-1} is the design weight assigned to household k (π_k being the sampling fraction).

2. Calculation of the slippage rate

The slippage rate for a control group c, written $rate_c$, is

$$rate_c = 100 * \frac{\left(\sum_{k \in s_{c,r}} w_k^{NR}\right) - t_c}{t_c}$$

where

$s_{c,r}$ is the set of respondent households in control group c,

w_k^{NR} is the sub-weight for household k (see definition of w_k^{NR} above), and

t_c is the total of the auxiliary data for control group c.

3. Calculation of the impact of imputation

The impact of imputation for an expenditure type a, written imp_e_a, is

$$imp_e_a = \frac{\sum_{k \in s_{i,r}} w_k^f * e_{a,k}}{\sum_{k \in s_{i,r}} w_k^f * e_{a,k} + \sum_{k \in s_{ni,r}} w_k^f * e_{a,k}}$$

where

$s_{i,r}$ is the set of respondent households having the type of expenditure a imputed,

$s_{ni,r}$ is the set of respondent households having the type of expenditure a not imputed,

w_k^f is the final weight assigned to household k, and

$e_{a,k}$ is the amount for the type of expenditure a for household k.

Appendix B
2001 Food Expenditure Survey Questionnaire (FE2)

Food Expenditure Survey in 2001

Questionnaire

FE2

Confidential when completed

Collected under the authority of the *Statistics Act*, Revised Statutes of Canada, 1985, Chapter S19.

Français au verso

001

Stratum	Type	Cluster	Rot	List

002 M | 003 Month | 004 N.I.

Record of drop-off and pick-ups

Visit	Day	Month	Began	Ended
Drop-off diaries	005	006	007 :	008 :
Pick up first diary	009	010	011 :	012 :
Pick up second diary	013	014	015 :	016 :

Record of Visits

Visit Number	Date	Time	Comments

Interviewer's Name *(Please print)*

Interviewer No.
045

For information only

8-5400-14 2000-07-14 STC/ISD 045-60119

Statistics Canada / Statistique Canada

Canada

Detach and destroy after processing

Listing Address

Household Contact

Telephone No.
(Area code)

A. Household composition

	001	011
1. What are the **first names** of all members of your household? *List the household reference person first (see definition). Ask all questions in Section A for each member of the household that you have listed.* **Household Reference Person**: The member of the household mainly responsible for its financial maintenance (i.e., pays the rent, mortgage, property taxes or electricity, etc.). This person can be either male or female. In cases where members share equal financial responsibility, choose one to be the household reference person.	**Person** 0 1 First Name	**Person** First Name
2. What is _ _____ 's **relationship** to the household reference person?	002 1 ⊗ Reference Person	012 2 ○ Spouse 3 ○ Son/Daughter 4 ○ Other relative 5 ○ Not related
3. In what month and year was _____ born? *(If born in 1900 or earlier, enter 1900.)*	003 M M Y Y Y Y	013 M M Y Y Y Y
4. Is _____ male or female?	004 1 ○ Male 2 ○ Female	014 1 ○ Male 2 ○ Female
5. What is _____ 's **marital status?** *Mark one circle.*	005 1 ○ Married spouse of a household member 2 ○ Common-law spouse of a household member 3 ○ Never married (single) 4 ○ Other (separated, divorced or widowed)	015 1 ○ Married spouse of a household member 2 ○ Common-law spouse of a household member 3 ○ Never married (single) 4 ○ Other (separated, divorced or widowed)
6. ***Economic Family Code*** **Economic Family**: Two or more persons who live in the same dwelling and are related to each other by blood, marriage, adoption or common-law. Unrelated roommates would have different codes.	006 *Enter Code* A	016 *Enter Code*

Check household membership

After listing members of the household ask:

- Does anyone else live at this address?
- Are there any persons away who could be considered as living at this address?

For information only

A. Household composition

	Person	Person	Person	Person
	First Name	First Name	First Name	First Name
022 / 032 / 042 / 052	1 Spouse 2 Son/Daughter 3 Other relative 4 Not related	1 Spouse 2 Son/Daughter 3 Other relative 4 Not related	1 Spouse 2 Son/Daughter 3 Other relative 4 Not related	1 Spouse 2 Son/Daughter 3 Other relative 4 Not related
023 / 033 / 043 / 053	M M Y Y Y Y	M M Y Y Y Y	M M Y Y Y Y	M M Y Y Y Y
024 / 034 / 044 / 054	1 Male 2 Female	1 Male 2 Female	1 Male 2 Female	1 Male 2 Female
025 / 035 / 045 / 055	1 Married spouse of a household member 2 Common-law spouse of a household member 3 Never married (single) 4 Other (separated, divorced or widowed)	1 Married spouse of a household member 2 Common-law spouse of a household member 3 Never married (single) 4 Other (separated, divorced or widowed)	1 Married spouse of a household member 2 Common-law spouse of a household member 3 Never married (single) 4 Other (separated, divorced or widowed)	1 Married spouse of a household member 2 Common-law spouse of a household member 3 Never married (single) 4 Other (separated, divorced or widowed)
026 / 036 / 046 / 056	Enter Code	Enter Code	Enter Code	Enter Code

Notes and Comments

For information only

B. Spending habits

Excluding purchases made while away from home overnight or longer, in the last four weeks . . .	**Total Cost**
1. How much do you estimate this household spent on food and other groceries purchased from stores (including farmer stalls and home delivery)? • **Exclude** bulk purchases of food *(See Q 3)*	001 $
2. How much of this amount was for **non-food items** such as paper products, household supplies, pet food, alcoholic beverages, etc.?	002 $
3. What amount was spent for bulk purchases of food, e.g. meat IN EXCESS of 25 kg (55 lbs.); bulk quantities of fruit or vegetables for canning, freezing, etc.? • **Include** charges for cutting, wrapping and freezing	003 $
4. Did this household buy any prepared food or non-alcoholic beverages from stores for parties, weddings and other occasions not reported in Q. 1 or Q. 3 above? • **Exclude** restaurants and caterers. 4.1 What amount was spent?	004 1 ◯ Yes → *Continue* 2 ◯ No → *Go to Q. 5* 005 $

5. Looking at the scale, what is your best estimate of the **total** income from all sources, **before** deductions, of **all** household members during the past **12 months**?

030
- 1 ◯ Less than $5,000
- 2 ◯ $5,000 - $9,999
- 3 ◯ $10,000 - $14,999
- 4 ◯ $15,000 - $19,999
- 5 ◯ $20,000 - $29,999
- 6 ◯ $30,000 - $39,999
- 7 ◯ $40,000 - $49,999
- 8 ◯ $50,000 - $59,999
- 9 ◯ $60,000 - $69,999
- 10 ◯ $70,000 - $79,999
- 11 ◯ $80,000 - $99,999
- 12 ◯ $100,000 or more
- 97 ◯ Don't know
- 98 ◯ Refused

Notes and Comments

For information only

C. Food and beverages while away from home overnight or longer during the previous month

1. Were any household members away from home overnight or longer during the month of _____ ?
(Interviewer: Indicate previous month.)

001 ○ Yes → *Continue*
○ No → *End*

2. For each absence from home overnight or longer during the previous month indicate the number of persons away and the number of nights per person.

	Absence No. 1	Absence No. 2	Absence No. 3	Absence No. 4	Absence No. 5
Number of persons away	002	004	006	008	010
Nights away	003	005	007	009	011

3. While away, how much board, if any, was paid to a private household by members of this household?

020 $

4. For all these absences, I need some information about the food and non-alcoholic beverages PURCHASED FROM RESTAURANTS.

- **Include** tips and taxes as well as purchases made for guests.
- **Exclude** expenses which were reimbursed or part of a package.

		Table service restaurants	Fast food restaurants	Cafeterias	Other
Breakfasts	Number of meals	021	023	025	027
	Expenditures	022 $	024 $	026 $	028 $
Lunches	Number of meals	029	031	033	035
	Expenditures	030 $	032 $	034 $	036 $
Dinners	Number of meals	037	039	041	043
	Expenditures	038 $	040 $	042 $	044 $
Between-meal food, snacks, non-alcoholic beverages	Expenditures	045 $	046 $	047 $	048 $

Note: "Other" restaurants include refreshment stands, snack bars, vending machines, mobile canteens, chip wagons, caterers, coffee wagons, etc.

5. How much did this household spend on food and non-alcoholic beverages PURCHASED FROM STORES while away from home overnight or longer during the previous month? For example, snack food purchased from gasoline stations, food purchased from grocery stores, etc.

050 $

6. How many meals were received free, reimbursed, or part of a package trip (e.g., business, dinner at friends)?

051

For information only

D. Diary follow-up report

Interviewer: Please complete after each diary pick-up.	Week 1	Week 2
1. a) During or after the follow-up, did you or the respondent enter any items in the FOOD FROM STORES section of the diary?	001 1 ◯ Yes, **all** of the items → ***Go to Q. 2*** 2 ◯ Yes, **some** items *Please estimate the value of these items.* $ [002] → ***Go to Q. 2*** 3 ◯ No → ***Continue***	009 1 ◯ Yes, **all** of the items → ***Go to Q. 2*** 2 ◯ Yes, **some** items *Please estimate the value of these items.* $ [010] → ***Go to Q. 2*** 3 ◯ No → ***Continue***
b) If nothing was entered, what was the reason?	003 1 ◯ The respondent had already recorded all purchases 2 ◯ The respondent did not make any purchases from stores 3 ◯ Other *Please comment in Q.5*	011 1 ◯ The respondent had already recorded all purchases 3 ◯ The respondent did not make any purchases from stores 4 ◯ Other *Please comment in Q.5*
2. a) During or after the follow-up, did you or the respondent enter any items in the FOOD FROM RESTAURANTS section of the diary?	004 1 ◯ Yes, **all** of the items → ***Go to Q. 3*** 2 ◯ Yes, **some** items *Please estimate the value of these items.* $ [005] → ***Go to Q. 3*** 3 ◯ No → ***Continue***	012 1 ◯ Yes, **all** of the items → ***Go to Q. 3*** 2 ◯ Yes, **some** items *Please estimate the value of these items.* $ [013] → ***Go to Q. 3*** 3 ◯ No → ***Continue***
b) If nothing was entered, what was the reason?	006 1 ◯ The respondent had already recorded all purchases 2 ◯ The respondent did not make any purchases from restaurants 3 ◯ Other *Please comment in Q.5*	014 1 ◯ The respondent had already recorded all purchases 2 ◯ The respondent did not make any purchases from restaurants 3 ◯ Other *Please comment in Q.5*
3. What is the final completion status of the FOOD FROM STORES section of the diary?	007 1 ◯ Complete 2 ◯ Incomplete	015 1 ◯ Complete 4 ◯ Incomplete
4. What is the final completion status of the FOOD FROM RESTAURANTS section of the diary?	008 3 ◯ Complete 4 ◯ Incomplete	016 3 ◯ Complete 4 ◯ Incomplete

For information only

5. Please comment on any sections that were incomplete, or any problems that were encountered in completing the diaries.

Week 1

Week 2

Notes and comments

For information only

Appendix C
2001 Food Expenditure Survey Diary (FE3) — Example for one day

Please unfold the flaps to refer to the notes

Attach receipts here

Day 1

Check (✓) the appropriate day:

☐ MON ☐ TUE ☐ WED ☐ THUR ☐ FRI ☐ SAT ☐ SUN

Food and beverages purchased from STORES

Check (✓) here if no purchases made this day ☐

Describe the item purchased • Give a **complete** description of each item of food bought • **Avoid** abbreviations and print clearly • **Include** bulk purchases	**Was this product . . . ?** **A** = Fresh **B** = Frozen **C** = Canned/ bottled/plastic container **D** = Dried **E** = Other Enter code letter	**Purchased at . . .** **A** = Supermarket **B** = Food specialty store or market **C** = Convenience store **D** = Other Enter code letter	**Number of . . .** Cans, bottles, cartons, bags, boxes, packages, bunches, heads, dozens, etc. Enter quantity	**Weight or volume per item** Examples: 1 2 kg 2 5 litres 3 6 lbs 4 12 oz head dozen bunch	**Total cost** **Exclude** all taxes $ c	Office use only
Example: *COLA*	*C*	*A*	*24*	*355 ML*	*6 99*	

For information only

6

Please unfold the flaps to refer to the notes

Day 1 *continued*

Food and beverages purchased from STORES

Describe the item purchased • Give a **complete** description of each item of food bought • **Avoid** abbreviations and print clearly • **Include** bulk purchases	**Was this product . . . ?** **A** = Fresh **B** = Frozen **C** = Canned, bottled/plastic container **D** = Dried **E** = Other Enter code letter	**Purchased at . . .** **A** = Supermarket **B** = Food specialty store or market **C** = Convenience store **D** = Other Enter code letter	**Number of . . .** Cans, bottles, cartons, bags, boxes, packages, bunches, heads, dozens, etc. Enter quantity	**Weight or volume per item** Example: 1.2 kg 2.5 litres 5.0 lbs 19 oz head dozen bunch	**Total cost** **Exclude** all taxes $ c	Office use only
Example: *GREEN ONIONS*	*A*	*B*	*1*	*BUNCH*	*79*	

Food and beverages purchased from RESTAURANTS

Check (•) here if no purchases made this day ☐

Whose meal or snack? **Include** • food from restaurants, cafeterias, take-outs, vending machines, canteens, etc. • snacks, non-alcoholic beverages, ice cream, candy, etc. • meals bought for guests **Exclude** alcoholic beverages	**Meal type:** **A** = Breakfast **B** = Lunch **C** = Dinner **D** = Between-meal snack and other food Enter code letter	**Number of meals purchased** Enter "0" if snacks or other food	**Restaurant type:** **A** = Table service **B** = Fast food **C** = Cafeteria **D** = Other Enter code letter	**Total cost** **Include** all taxes and tips on food and non-alcoholic beverages **Exclude** taxes and tips on alcoholic beverages $ c
Example: *TAYLOR AND SONS LUNCH*	*B*	*2*	*B*	*8 69*

If there is insufficient space to enter your purchases made this day, use pages 20 to 23, "ADDITIONAL PAGE".

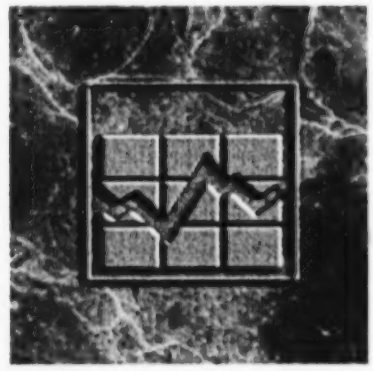
For information only

N° 62F0026MIF au catalogue — N° 003
ISSN : 1708-8887
ISBN : 0-662-74537-X

Document de recherche

Série de documents de recherche sur les dépenses des ménages

Enquête sur les dépenses alimentaires de 2001
Indicateurs de qualité des données



par Sylvie Auger, Khushnood Khan, Sylvain Nadon et Johanne Tremblay

Division des méthodes d'enquêtes auprès des ménages
Immeuble R.H. Coats, Ottawa, K1A 0T6

Téléphone: 613 951-7355

Statistique Canada Statistics Canada

Canada

Comment obtenir d'autres renseignements

Toute demande de renseignements au sujet du présent produit ou au sujet de statistiques ou de services connexes doit être adressée à : Division de la statistique du revenu, Statistique Canada, Ottawa, Ontario, K1A 0T6 (613) 951-7355; (888) 297-7355; revenu@statcan.ca).

Pour obtenir des renseignements sur l'ensemble des données de Statistique Canada qui sont disponibles, veuillez composer l'un des numéros sans frais suivants. Vous pouvez également communiquer avec nous par courriel ou visiter notre site Web.

Service national de renseignements	1 800 263-1136
Service national d'appareils de télécommunications pour les malentendants	1 800 363-7629
Renseignements concernant le Programme des services de dépôt	1 800 700-1033
Télécopieur pour le Programme des services de dépôt	1 800 889-9734
Renseignements par courriel	infostats@statcan.ca
Site Web	www.statcan.ca

Renseignements pour accéder au produit

Le produit nº 62F0026MIF au catalogue est disponible gratuitement. Pour obtenir un exemplaire, il suffit de visiter notre site Web à www.statcan.ca et de choisir la rubrique Nos produits et services.

Normes de service à la clientèle

Statistique Canada s'engage à fournir à ses clients des services rapides, fiables et courtois, et ce, dans la langue officielle de leur choix. À cet égard, notre organisme s'est doté de normes de service à la clientèle qui doivent être observées par les employés lorsqu'ils offrent des services à la clientèle. Pour obtenir une copie de ces normes de service, veuillez communiquer avec Statistique Canada au numéro sans frais 1 800 263-1136. Les normes de service sont aussi publiées dans le site www.statcan.ca sous À propos de Statistique Canada > Offrir des services aux Canadiens.

Statistique Canada
Division de la statistique du revenu

Série de documents de recherche sur les dépenses des ménages

Enquête sur les dépenses alimentaires de 2001
Indicateurs de qualité des données

Publication autorisée par le ministre responsable de Statistique Canada

© Ministre de l'Industrie, 2005

Tous droits réservés. Le contenu de la présente publication peut être reproduit, en tout ou en partie, et par quelque moyen que ce soit, sans autre permission de Statistique Canada sous réserve que la reproduction soit effectuée uniquement à des fins d'étude privée, de recherche, de critique, de compte rendu ou en vue d'en préparer un résumé destiné aux journaux, et/ou à des fins non commerciales. Statistique Canada doit être cité comme suit : Source (ou « Adapté de », s'il y a lieu) : Statistique Canada, nom du produit, numéro au catalogue, volume et numéro, période de référence et page(s). Autrement, il est interdit de reproduire quelque contenu de la présente publication, ou de l'emmagasiner dans un système de recouvrement, ou de le transmettre sous quelque forme et par quelque moyen que ce soit, reproduction électronique, mécanique, photographique, pour quelque fin que ce soit, sans l'autorisation écrite préalable des Services d'octroi de licences, Division du marketing, Statistique Canada, Ottawa, Ontario, Canada K1A 0T6.

Juillet 2005

N° 62F0026MIF2005003 au catalogue

ISSN : 1708-8887
ISBN : 0-662-74537-X

Périodicité : hors série

Ottawa

This publication is available in English upon request (Catalogue no. 62F0026MIE).

Note de reconnaissance

Le succès du système statistique du Canada repose sur un partenariat bien établi entre Statistique Canada et la population, les entreprises, les administrations canadiennes et les autres organismes. Sans cette collaboration et cette bonne volonté, il serait impossible de produire des statistiques précises et actuelles.

Résumé

L'Enquête sur les dépenses alimentaires (EDA) est une enquête périodique recueillant des informations auprès des ménages sur les habitudes de dépense alimentaire. La collecte des renseignements se fait principalement à partir de journaux hebdomadaires des achats que les répondants doivent remplir quotidiennement durant deux semaines consécutives.

L'EDA, comme toutes les enquêtes, est sujette aux erreurs et ce malgré toutes les précautions prises aux différentes étapes de l'enquête pour les contrôler. Quoiqu'il n'existe pas de mesure exhaustive de la qualité des données d'une enquête, certaines mesures de qualité produites aux différentes étapes de l'enquête peuvent fournir à l'utilisateur de l'information pertinente pour une bonne interprétation des données.

Ce document présente, pour l'EDA de 2001, les indicateurs de qualité suivants : les coefficients de variation, les taux de non-réponse, les taux de vacance, les taux de glissement, les taux d'imputation ainsi que les incidences de l'imputation sur les estimations.

Table des matières

Faits saillants 6

Introduction 8

1. Les erreurs d'échantillonnage 9

1.1 Les mesures de l'erreur d'échantillonnage 9
1.2 Les coefficients de variation 9
1.3 Modèle pour dériver une approximation du CV 11
1.4 Suppression des données non fiables dans les tableaux d'estimations 12

2. La non-réponse 13

2.1 Les taux de non-réponse et les taux de vacance 14
2.2 La non-réponse selon le niveau d'urbanisation 16
2.3 La non-réponse selon les strates de revenu 18
2.4 L'ajustement pour la non-réponse 19

3. Erreurs de couverture 19

3.1 Le sous-dénombrement et le surdénombrement : les taux de glissement 20
3.2 L'ajustement pour l'erreur de couverture au niveau de la population et des ménages 22

4. Les erreurs de réponse 23

5. Les erreurs de traitement 24

5.1 Données sur l'imputation 26

Bibliographie 33

Annexe A 34

Notation algébrique

Annexe B 36

Questionnaire de l'Enquête sur les dépenses alimentaires de 2001 (FE2)

Annexe C 43

Journal de l'Enquête sur les dépenses alimentaires de 2001 (FE3) — Exemple d'une journée

Faits saillants

Les erreurs d'échantillonnage

- Les coefficients de variation (CV) de l'estimation du total des dépenses hebdomadaires moyennes par ménage varient de 1,9 % à 3,1 % à l'échelle régionale (cinq grandes régions canadiennes) et le CV est de 1,1 % à l'échelle nationale.

- Les coefficients de variation des dépenses hebdomadaires moyennes par ménage selon les principales catégories de dépenses sommaires sont généralement inférieurs à 4 % à l'échelle nationale et inférieurs à 9 % à l'échelle régionale.

La non-réponse

- Le taux de non-réponse est de 28,5 %. La non-réponse est due aux refus (17,0 %) et aux ménages qu'on n'a pas réussi à contacter (11,4 %).

- Le taux de non-réponse final a tendance à croître avec le niveau d'urbanisation. On observe un taux de non-réponse de 19,2 % en milieu rural, et un taux de 30,0 % dans les centres urbains d'un million d'habitants et plus.

- L'analyse des taux de réponse finaux dans les strates d'aires géographiques à revenu élevé et à faible revenu provenant du plan d'échantillonnage révèle que le taux de non-réponse dans les strates à revenu élevé (37,2 %) est plus élevé que celui observé dans les strates à faible revenu (30,4 %) et dans les strates régulières (27,8 %).

Les erreurs de couverture

- On observe un sous-dénombrement des ménages de 7,9 % à l'échelle nationale. On observe aussi un sous-dénombrement des ménages pour chacune des régions, les taux variant de 5,0 % à 10,4 %.

- On observe un sous-dénombrement des personnes de 6,9 % à l'échelle nationale. On observe aussi un sous-dénombrement des personnes pour chacune des régions, les taux variant de 4,6 % à 9,2 %.

- Les taux de glissement nationaux pour les enfants (0 à 6 ans et 7 à 17 ans) sont très différents de ceux des autres groupes d'âge. En effet, on observe un surdénombrement chez les enfants et un sous-dénombrement chez les adultes.

Les erreurs de réponse

- La collecte des renseignements se fait par une entrevue en personne au moyen d'un questionnaire papier, ainsi qu'avec deux journaux hebdomadaires des achats que le répondant doit remplir quotidiennement durant deux semaines consécutives. L'estimation des dépenses obtenue par l'entremise du questionnaire est de 8,3 % supérieure à celle obtenue à partir des journaux.

Les erreurs de traitement liées à l'imputation

i) Dans le cas du questionnaire (FE2)

- Deux sections seulement sont imputées : celle des *Habitudes de dépense du ménage* (section B) et celle des *Aliments et boissons consommés hors du lieu de résidence – pendant une absence d'une nuit ou plus au cours du mois précédent* (section C). Dans le cas de la section B, une proportion de 12,1 % des questionnaires est imputée. Dans la plupart d'entre eux, une ou deux variables sur cinq doivent être imputées. Pour ce qui est de la section C, 4,2 % des questionnaires sont imputés.

ii) Dans le cas du journal (FE3)

- Pour chaque jour, seule la section des *Aliments et boissons achetés dans un magasin* est imputée. Des 279 709 dépenses alimentaires déclarées, 2,2 % sont imputées. La plupart du temps, l'imputation précise le type de dépense. Par exemple, le répondant indique qu'il a acheté du lait sans spécifier le type de lait (lait entier de consommation, lait à faible teneur en gras, lait écrémé, etc.).

Introduction

L'Enquête sur les dépenses alimentaires (EDA) est une enquête périodique qui a lieu à tous les 4 à 6 ans. Cette enquête recueille des informations auprès des ménages canadiens sur les habitudes de dépense alimentaire. La collecte des renseignements se fait par une entrevue en personne au moyen d'un questionnaire papier (FE2), ainsi qu'avec deux journaux hebdomadaires des achats (FE3) que le répondant doit remplir quotidiennement durant deux semaines consécutives. L'Enquête sur les dépenses alimentaires 2001 s'est déroulée tous les mois au cours de l'année civile 2001. L'échantillon de l'EDA 2001 est composé de 8 414 ménages répartis dans les dix provinces ainsi que dans les villes de Whitehorse, Yellowknife et Iqaluit.

Cette enquête sert à compléter l'Enquête sur les dépenses des ménages (EDM)[1] puisqu'elle procure des renseignements détaillés sur les dépenses alimentaires qu'il n'est pas possible d'obtenir avec la méthodologie utilisée pour l'EDM. L'Enquête sur les dépenses alimentaires et l'Enquête sur les dépenses des ménages servent à mettre à jour les pondérations utilisées dans l'Indice des prix à la consommation.

L'EDA, comme toutes les enquêtes, est sujette aux erreurs et ce malgré toutes les précautions prises aux différentes étapes de l'enquête pour les contrôler. Quoiqu'il n'existe pas de mesure exhaustive de la qualité des données d'une enquête, certaines mesures de qualité produites aux différentes étapes de l'enquête peuvent fournir à l'utilisateur de l'information pertinente pour une bonne interprétation des données.

Ce document vise donc à présenter les indicateurs de qualité produits pour l'Enquête sur les dépenses alimentaires de 2001. On y trouve les indicateurs de qualité usuels généralement utiles aux utilisateurs pour l'interprétation des données tels que les coefficients de variation, les taux de non-réponse, les taux de glissement et les taux d'imputation.

Les indicateurs de qualité ont été classés selon les principaux types d'erreur que l'on retrouve dans une enquête. Les erreurs d'échantillonnage, c'est-à-dire les erreurs dues au fait que les inférences tirées au sujet de l'ensemble de la population sont fondées sur des renseignements recueillis auprès d'un échantillon de la population et non pas auprès de toute la population, sont présentées dans la section 1. Les sections suivantes couvrent les erreurs non dues à l'échantillonnage. La non-réponse et les erreurs de couverture sont d'abord discutées dans les sections 2 et 3. Les erreurs de réponse et les erreurs de traitement sont traitées dans les sections 4 et 5 respectivement.

Ce document est axé sur la qualité des données. Pour une description détaillée de la méthodologie de l'enquête, on peut consulter la référence [1].

1. L'EDM est une enquête annuelle où l'on demande à un échantillon de ménages canadiens de faire état de toutes les dépenses effectuées au cours d'une année civile. Les résultats de l'EDM sont publiés dans Les habitudes de dépense au Canada, N° 62-202 au catalogue.

1. Les erreurs d'échantillonnage

Les erreurs d'échantillonnage découlent du fait que les inférences que l'on tire au sujet de l'ensemble de la population à partir de l'enquête sont fondées sur l'information que l'on a recueillie auprès d'un échantillon de la population et non pas auprès de toute la population. Outre le plan d'échantillonnage et la méthode d'estimation qui ont été appliqués pour l'Enquête sur les dépenses alimentaires, la taille d'échantillon ainsi que la variabilité de chaque caractéristique sont des facteurs déterminants de l'erreur d'échantillonnage. Les caractéristiques qui sont rares ou qui sont distribuées avec une grande variabilité dans la population auront une erreur d'échantillonnage plus grande que les caractéristiques qu'on observe plus fréquemment ou qui sont distribuées avec moins de variabilité dans la population.

1.1 Les mesures de l'erreur d'échantillonnage

L'erreur-type est une mesure standard de l'erreur d'échantillonnage. L'erreur-type correspond au degré de variation de l'estimation considérant qu'un échantillon particulier a été choisi plutôt qu'un autre parmi tous les échantillons possibles de même taille sous le même plan d'échantillonnage. Comme l'EDA utilise un plan d'échantillonnage et une méthode d'estimation complexes, on estime l'erreur-type à partir d'une méthode de rééchantillonnage appelée le Jackknife. Pour plus de détails sur cette méthode, voir la référence [2].

Le coefficient de variation (CV) est également une mesure de fiabilité de l'estimation fréquemment utilisée. Il exprime tout simplement l'erreur-type en pourcentage de l'estimation. Ainsi, si on obtient une estimation Y pour une certaine caractéristique et que SE correspond à l'erreur-type estimée, le CV sera (SE/Y) x 100.

Finalement, l'erreur-type ou le coefficient de variation peuvent être utilisés pour dériver une autre mesure de précision des estimations, l'intervalle de confiance. Cette mesure indique le niveau de confiance selon lequel la valeur réelle dans la population pour une caractéristique qu'on a observée se trouve à l'intérieur de certaines limites. Un intervalle avec un niveau de confiance de 95 % correspond à l'estimation obtenue à partir de l'échantillon $\pm$ 2 erreur-type, soit (Y $\pm$ 2 SE). Cela signifie que si on répétait l'échantillonnage un grand nombre de fois, chaque échantillon fournirait un intervalle différent et 95 % des intervalles contiendraient la vraie valeur de la caractéristique. De façon similaire, si on répétait l'échantillonnage, l'intervalle Y$\pm$ SE contiendrait la vraie valeur dans 68 % des cas. À noter que l'intervalle de confiance se calcule également à partir du CV de façon similaire, soit (Y $\pm$ 2 (CV x Y) / 100).

1.2 Les coefficients de variation

Des estimations des coefficients de variation sont calculées pour les estimations de plusieurs caractéristiques recueillies dans l'EDA. Les CV pour les estimations hebdomadaires de dépense moyenne par ménage, de pourcentage déclarant (proportion en pourcentage de journaux contenant une dépense particulière) et de quantité moyenne par ménage sont disponibles à l'échelle nationale et régionale dans la publication *Dépenses alimentaires au Canada*, voir la référence [3].

On doit noter que les estimations de CV ne tiennent pas compte du fait que certaines données ont été imputées; par conséquent, les CV présentés peuvent sous-estimer les valeurs réelles. Pour la plupart des variables, le taux d'imputation est peu élevé (voir la section 5) et les CV fournis représentent une bonne estimation des vrais CV. Cependant, il est important de tenir compte à la fois du CV et du taux d'imputation si on s'intéresse à la fiabilité des dépenses alimentaires détaillées ayant un taux d'imputation élevé.

Le tableau 1.1 donne un aperçu des CV des estimations des dépenses hebdomadaires moyennes par ménage à l'échelle régionale ainsi qu'à l'échelle nationale pour quelques-unes des catégories de dépenses sommaires.

Tableau 1.1
Coefficients de variation (%) des dépenses hebdomadaires moyennes par ménage pour quelques catégories de dépenses sommaires, cinq grandes régions canadiennes et Canada (10 provinces)

Catégories de dépenses sommaires	**Canada (10 prov)**	Atlantique	Québec	Ontario	Prairies	Colombie-Britannique
Aliments achetés au magasin	1,0	3,1	1,8	1,9	3,1	2,1
Au cours de voyages de plus d'un jour	10,5	22,4	13,6	13,7	35,4	15,3
Localement ou au cours de voyages d'un jour	1,0	2,9	1,8	1,8	2,3	2,1
Viande	1,5	4,5	2,6	2,6	3,9	3,5
Poisson et autres produits marins	3,5	8,2	7,2	6,0	7,3	8,5
Produits laitiers et œufs	1,2	3,1	2,4	2,3	2,6	2,7
Produits de boulangerie et autres produits céréaliers	1,2	3,0	2,0	2,3	2,5	2,6
Fruits et noix	1,4	4,4	2,6	2,5	3,0	3,2
Légumes	1,5	4,0	2,6	2,8	3,3	3,3
Condiments, épices et vinaigre	1,9	5,5	3,6	3,6	4,3	4,5
Sucre et préparations à base de sucre	2,3	5,6	5,0	4,2	4,5	5,4
Café et thé	2,9	7,1	5,4	5,0	8,3	6,4
Graisses et huiles	2,7	6,7	5,2	5,2	5,6	7,4
Autres aliments, ingrédients et préparations alimentaires	1,7	5,1	3,3	3,2	4,2	3,7
Boissons non alcoolisées	2,0	5,2	3,8	3,6	4,9	5,3
Aliments achetés au restaurant	2,2	5,9	4,4	4,4	4,4	4,5
Total des dépenses alimentaires hebdomadaires	1,1	2,9	1,9	2,0	3,1	2,1

Les CV de l'estimation du total des dépenses hebdomadaires moyennes par ménage varient de 1,9 % à 3,1 % à l'échelle régionale et le CV est de 1,1 % à l'échelle nationale. La catégorie des aliments achetés au magasin a des CV semblables à ceux du total, par contre les CV des aliments achetés au restaurant sont presque deux fois plus élevés.

On constate que les CV des aliments achetés au magasin au cours de voyages de plus d'un jour sont beaucoup plus élevés que les autres. À l'échelle nationale, le CV est de 10,5 % et à l'échelle régionale, ils varient de 13,6 % à 35,4 %.

La qualité des estimations varie selon les catégories d'aliments. Les CV des catégories « Poisson et autres produits marins », « Café et thé » et « Graisses et huiles » se situent autour de 3 % à l'échelle nationale et varient de 5 % à 9 % à l'échelle régionale. Les neuf autres catégories ont des CV variant principalement de 1 % à 2 % à l'échelle nationale et de 2 % à 6 % à l'échelle régionale.

1.3 Modèle pour dériver une approximation du CV

Des estimations pour différents domaines d'intérêt (par ex., par taille de la région de résidence) sont disponibles pour les catégories de dépenses sommaires dans la publication *Dépenses alimentaires au Canada* (référence [3]). Des estimations pour différents domaines d'intérêt pour les catégories de dépenses détaillées sont également disponibles sur demande auprès de la Division de la statistique du revenu. Pour des raisons opérationnelles, il n'est pas possible de produire les CV pour tous les différents niveaux d'agrégation qui pourraient intéresser les utilisateurs.

1.3.1 Approximation du CV pour les estimations des domaines

Il est toutefois possible de calculer une approximation du CV en utilisant une relation entre le nombre de journaux hebdomadaires dans lesquels des dépenses pour un article ont été déclarées et le CV à un niveau agrégé. Cette relation, basée sur la tendance du CV à croître proportionnellement à une diminution de la racine carrée du nombre de journaux hebdomadaires ayant un montant différent de zéro est illustrée ci-après.

Formule d'approximation du CV pour un domaine (un sous-groupe de la population)

Si CV(Y) représente le CV pour l'estimation de la dépense hebdomadaire moyenne par ménage d'une certaine caractéristique pour toute la population, alors on peut calculer une approximation du CV de l'estimation de cette caractéristique pour un domaine (que l'on peut considérer comme un sous-groupe de la population tel qu'un type de ménage, un niveau d'urbanisation, ...) à partir de l'équation suivante :

$$CV\ (Y_d) = CV\ (Y) \times \sqrt{\frac{n}{n_d}}$$

où

n = *le nombre de journaux hebdomadaires ayant un montant différent de zéro pour cette caractéristique*

qui peut être obtenu en calculant

$$\frac{\textit{le nombre de journaux hebdomadaires} \times \textit{l'estimation du pourcentage déclarant}^2}{100}$$

2. Proportion en pourcentage de journaux contenant un montant différent de zéro pour cette caractéristique.

n_d = *le nombre de journaux hebdomadaires ayant un montant différent de zéro pour cette caractéristique dans le domaine d*

qui peut être obtenu en calculant

$$\frac{\textit{le nombre de journaux hebdomadaires dans le domaine d} \times \textit{l'estimation du pourcentage déclarant}^2 \textit{ dans le domaine d}}{100}$$

Généralement on utilise le CV, le nombre de journaux hebdomadaires et le pourcentage déclarant à l'échelle nationale pour calculer les approximations pour les différents domaines. Dans le cas où on cherche à calculer une approximation du CV pour un domaine entièrement contenu dans une seule région (par ex. région métropolitaine), il est préférable d'utiliser ces valeurs à l'échelle régionale puisque les CV régionaux sont publiés pour l'EDA de 2001 (référence [3]). Il est important de se rappeler que la valeur du CV obtenue par cette approche n'est qu'une approximation du CV.

1.3.2 Approximation du CV à partir des fichiers de microdonnées

Les utilisateurs des fichiers de microdonnées peuvent se servir d'une autre approche pour dériver une approximation du CV des estimations, qui sera généralement plus performante que celle présentée dans la section précédente pour les CV des catégories de dépenses détaillées. Cette approche est décrite en détails dans la documentation accompagnant les fichiers de microdonnées de 2001. Elle est utilisable seulement à partir des fichiers de microdonnées puisqu'il est nécessaire d'avoir les données et les poids de chaque ménage pour calculer cette approximation.

Le document sur la qualité des données de l'Enquête sur les dépenses des ménages de 1997 (référence [4]) renferme des résultats de l'évaluation de la performance de ces deux méthodes d'approximation du CV.

1.4 Suppression des données non fiables dans les tableaux d'estimations

Comme le coefficient de variation est un indicateur de la fiabilité des données, on aimerait l'utiliser pour déterminer si les estimations devraient être publiées ou non. On considère que les estimations dont le CV est estimé à plus de 33 % ne sont pas suffisamment fiables pour être publiées. Afin de faciliter la mise en tableaux, la règle de suppression pour les estimations de dépenses alimentaires est basée sur le nombre de journaux hebdomadaires ayant un montant différent de zéro pour une catégorie de dépenses.

Selon l'analyse d'enquêtes antérieures, les CV sont en général inférieurs à 33 % si le nombre de journaux hebdomadaires ayant un montant différent de zéro pour un article est supérieur à 40. Comme il s'agit d'une règle approximative, certaines estimations peuvent être publiées même si le CV est supérieur à 33 % (erreur de type 1) et certaines estimations ne seront pas publiées malgré un CV inférieur à 33 % (erreur de type 2). Le tableau 1.2 ci-dessous indique la place qu'occupent ces deux types d'erreur pour les CV

d'estimations de dépenses hebdomadaires moyennes par ménage de 365 articles alimentaires.

Tableau 1.2
Erreur de type 1 et 2 (%) selon la règle de 40 journaux

Domaine	CV = 0	Erreur de type 1	Erreur de type 2	Nombre de CV impliqués
Canada (10 provinces)	0,0	1,6	0,3	365
Région	0,5	0,6	7,7	1 825
Province	2,9	0,7	14,8	3 650
Région métropolitaine	7,0	0,7	18,3	5 840
Niveau d'urbanisation	0,8	0,9	10,2	2 555
Taille de ménage	0,5	0,5	6,8	1 825
Composition du ménage	3,2	0,9	12,3	2 920

La colonne « CV = 0 » représente la proportion d'estimations dont les CV sont égaux à zéro. Ils sont dus principalement en raison qu'il n'y a aucun journal hebdomadaire ayant un montant différent de zéro pour les estimations associées.

La colonne « Erreur de type 1 » représente la proportion d'estimations, qui seraient publiées selon la règle de 40 journaux, dont les CV sont de plus de 33 %. Cette erreur est de 1,6 % au niveau du Canada et moins de 1 % pour les autres domaines étudiés[3].

La colonne « Erreur de type 2 » représente la proportion d'estimations, qui ne seraient pas publiées selon la règle de 40 journaux, dont les CV sont de 33 % ou moins.

2. La non-réponse

Les erreurs dues à la non-réponse découlent du fait que certains répondants potentiels ne fournissent pas l'information nécessaire ou que cette information s'avère inutilisable. Lorsque le répondant a omis de répondre à certaines questions seulement, on parle alors de non-réponse partielle. Dans ce cas, les données manquantes sont imputées. Les erreurs liées à l'imputation sont présentées dans la section 5 portant sur les erreurs de traitement.

Il y a également un autre type de non-réponse partielle : les répondants qui n'ont rapporté qu'une semaine de données. Au niveau des dix provinces, il y en a 252.

3. Cette erreur est plus élevée au niveau du Canada en raison de la proportion d'estimations publiées par rapport aux autres domaines. Cette proportion est plus de 98 % au niveau du Canada et moins de 85 % pour les autres domaines.

Dans la présente section, la non-réponse couvre la non-réponse à la collecte, due principalement à l'impossibilité de contacter le ménage ou au refus des membres du ménage de participer à l'enquête.

Le principal impact de la non-réponse sur la qualité des données est qu'elle peut induire un biais dans les estimations si les caractéristiques des répondants et des non-répondants diffèrent et que cette différence a un impact sur les caractéristiques étudiées. Les taux de non-réponse peuvent être calculés facilement mais ils n'ont qu'une valeur indicative à l'égard de la qualité des données car ils ne permettent pas de mesurer l'importance du biais associé aux estimations. L'ampleur de la non-réponse peut être considérée comme une évaluation des risques de biais dans les estimations.

2.1 Les taux de non-réponse et les taux de vacance

Dans l'EDA, comme les unités sélectionnées sont des logements, les intervieweurs doivent d'abord identifier les logements inadmissibles, c'est-à-dire les logements occupés par des personnes ne faisant pas partie de la population cible, les logements qui n'existent plus (démoli, maison mobile déplacée ou logement converti en entreprise) et les logements vacants (inoccupés, saisonniers ou en construction).

Parmi les logements admissibles, on évalue ensuite la proportion des ménages qui n'ont pas répondu à l'enquête, qu'on appelle le taux de non-réponse à la collecte. Ces derniers comprennent les ménages qui ont refusé de participer à l'enquête et les ménages où il a été impossible d'établir un contact avec les répondants parce qu'ils étaient absents ou encore à cause de circonstances spéciales (problème de langue, maladie ou décès).

Le tableau 2.1 présente le taux de non-réponse ventilé selon les refus et les non-contacts. On y trouve également le taux de vacance. Ces taux sont non pondérés. Ils sont fournis à l'échelle nationale ainsi qu'à l'échelle régionale et pour les trois villes du Nord, par trimestre et pour l'année.

Tableau 2.1
Taux de non-réponse (%) et taux de vacance (%) par trimestre et pour l'année, cinq grandes régions canadiennes, Canada (10 provinces) et les trois grandes villes des Territoires

Trimestres		Canada (10 prov)	Atl.	Qc	Ont.	Pr.	C.-B.	3 grandes villes des Territoires	Whitehorse	Yellowknife	Iqaluit
Janvier à décembre	Vacants	7,5	12,4	7,4	6,2	8,4	5,8	13,9	16,8	10,7	14,2
	Non-réponse	28,5	20,4	22,8	33,2	23,6	37,0	32,4	41,1	18,2	39,8
	Pas de contact	11,4	8,2	7,2	14,5	9,2	15,3	14,6	17,3	8,3	19,5
	Refus	17,0	12,3	15,6	18,7	14,3	21,7	17,8	23,8	9,9	20,3
Janvier à mars	Vacants	8,2	13,1	8,4	7,1	8,2	5,8	18,4	28,8	18,6	4,1
	Non-réponse	34,7	22,4	29,6	42,1	28,4	45,1	37,8	25,5	31,4	57,8
	Pas de contact	15,0	9,3	10,4	20,4	11,4	19,8	17,5	12,8	7,8	33,3
	Refus	19,7	13,1	19,2	21,7	17,0	25,3	20,3	12,8	23,5	24,4
Avril à juin	Vacants	7,7	9,2	7,7	6,0	10,2	7,1	12,8	9,8	13,3	16,3
	Non-réponse	29,7	20,1	23,6	32,9	30,0	36,8	39,7	55,8	26,3	29,0
	Pas de contact	11,3	6,4	7,8	13,7	10,8	14,7	18,2	21,2	15,8	16,1
	Refus	18,5	13,7	15,8	19,3	19,3	22,2	21,5	34,6	10,5	12,9
Juillet à septembre	Vacants	7,3	11,5	6,5	6,9	8,1	5,3	8,5	9,5	3,3	14,3
	Non-réponse	27,0	17,3	18,8	31,1	28,1	33,7	28,9	42,1	15,1	28,1
	Pas de contact	10,5	6,1	5,3	12,3	12,4	13,8	13,4	19,3	9,4	9,4
	Refus	16,5	11,2	13,5	18,9	15,7	19,9	15,5	22,8	5,7	18,8
Octobre à décembre	Vacants	6,9	15,3	7,0	4,8	7,2	5,0	15,2	18,3	6,5	23,8
	Non-réponse	21,3	21,4	18,7	25,9	6,1	31,8	23,1	39,1	2,0	36,0
	Pas de contact	8,3	10,7	5,1	11,1	1,7	12,7	9,1	15,2	2,0	12,0
	Refus	13,0	10,7	13,6	14,8	4,4	19,1	14,0	23,9	0,0	24,0

Le taux de non-réponse pour l'année au Canada (10 provinces) est de 28,5 %. Il est dû aux refus (17,0 %) et aux ménages qu'il a été impossible de rejoindre (11,4 %). Pour toutes les régions autres que le Nord, peu importe le trimestre, les refus représentent la première cause de non-réponse. Les trois villes du Nord réunies ont un taux de non-réponse pour l'année de 32,4 %. Il est dû aux refus (17,8 %) et aux ménages qu'il a été impossible de rejoindre (14,6 %). Dans les cas de Yellowknife et Iqaluit, pour certains trimestres la cause première de non-réponse est les non-contacts.

Les taux de non-réponse ont tendance à diminuer d'un trimestre à l'autre. On remarque que ces taux au premier trimestre sont particulièrement élevés en Ontario et en Colombie-Britannique. Ceux-ci reflètent des problèmes de collecte rencontrés au premier trimestre.

Les régions de l'Ontario et de la Colombie-Britannique ont les taux de non-réponse les plus élevés. Au niveau de l'année leurs taux de non-réponse sont respectivement 33,2 % et 37,0 % tandis que les trois autres régions ont des taux de non-réponse entre 20,4 % et 23,6 %.

Parmi les trois villes du Nord, Yellowknife est celle ayant le plus faible taux de non-réponse avec 18,2 % pour l'année. Whitehorse et Iqaluit ont des taux de non-réponse pour l'année de 41,1 % et 39,8 % respectivement.

Les taux de vacance sont présentés au tableau 2.1, mais on doit considérer que les logements vacants ne contribuent pas au biais de l'échantillon dans la mesure où ils sont identifiés correctement. En analysant les taux de vacance, on peut déceler les problèmes d'identification des logements liés à la collecte. Le taux de vacance de l'EDA de 2001 est de 7,5 % au Canada (10 provinces) et de 13,9 % pour l'ensemble des trois villes du Nord.

2.2 La non-réponse selon le niveau d'urbanisation

La non-réponse varie selon le niveau d'urbanisation. Les divers taux à l'échelle du Canada (10 provinces) sont présentés par niveau d'urbanisation, par trimestre et pour l'année dans le tableau 2.2.

Tableau 2.2
Taux de non-réponse (%) et taux de vacance (%) selon le niveau d'urbanisation, par trimestre et pour l'année, Canada (10 provinces)

Catégorie d'urbanisation		TOTAL	URBAIN	1 000 000 ou plus	500 000 à 999 999	250 000 à 499 999	100 000 à 249 999	30 000 à 99 999	Moins de 30 000	RURAL
Janvier à décembre	Vacants	7,5	4,4	3,4	3,4	5,1	5,6	5,0	6,6	22,6
	Non-réponse	26,5	30,0	36,4	28,5	28,9	22,2	27,9	20,9	19,2
	Pas de contact	11,4	12,2	16,1	11,3	10,8	6,7	10,1	8,0	7,0
	Refus	17,0	17,8	20,3	17,2	18,0	15,4	17,8	12,8	12,2
Janvier à mars	Vacants	8,2	4,8	4,0	3,6	8,0	5,2	5,4	5,2	23,8
	Non-réponse	34,7	36,5	45,0	38,2	35,3	24,6	30,5	26,0	24,3
	Pas de contact	15,0	15,8	22,0	17,5	16,3	6,1	9,3	9,2	9,8
	Refus	19,7	20,6	23,0	20,7	18,9	18,4	21,2	16,8	14,5
Avril à juin	Vacants	7,7	4,8	4,3	4,1	4,6	8,4	2,1	6,2	21,1
	Non-réponse	29,7	31,1	36,3	31,3	23,9	28,5	31,6	21,3	22,0
	Pas de contact	11,3	12,1	16,3	10,8	6,1	8,2	11,8	7,9	6,6
	Refus	18,5	19,0	19,9	20,5	17,8	20,3	19,9	13,4	15,4
Juillet à septembre	Vacants	7,3	4,6	2,9	3,1	5,1	5,3	6,8	10,0	20,5
	Non-réponse	27,0	28,6	31,8	29,0	29,9	19,3	31,4	23,2	16,9
	Pas de contact	10,5	11,1	13,1	11,2	9,1	6,8	14,0	8,2	6,5
	Refus	16,5	17,5	18,8	17,8	20,7	12,5	17,4	14,9	10,4
Octobre à décembre	Vacants	6,9	3,3	2,4	2,7	2,2	4,1	5,9	5,5	24,5
	Non-réponse	21,3	22,8	32,0	13,0	25,6	16,4	17,5	11,5	11,9
	Pas de contact	8,3	9,0	12,6	4,0	10,8	6,2	5,6	6,5	4,2
	Refus	13,0	13,8	19,4	9,1	14,8	10,2	11,9	5,0	7,7

Le taux de non-réponse pour l'année a tendance à augmenter avec le niveau d'urbanisation. Il y a un écart de 17,2 % entre les catégories 1 000 000 ou plus (36,4 %) et Rural (19,2 %). Les refus représentent la cause première de non-réponse à l'exception d'un cas : les moins de 30 000 dans le dernier trimestre.

En examinant les taux de vacance par niveau d'urbanisation, il ressort que le taux de vacance est plus élevé en région rurale (22,6 %) que dans les régions urbaines à faible population (6,6 %). Ce phénomène s'observe également dans l'Enquête sur la population active (EPA) et s'explique sans doute par un plus grand nombre de logements saisonniers en milieu rural. Comme l'échantillon de l'EDA est plus concentré dans les régions urbaines à forte population que celui de l'EPA, on peut s'attendre à observer un taux de vacance national légèrement inférieur à celui de l'EPA. Effectivement, l'EDA 2001 a un taux de vacance national de 7,5 % tandis que le taux de

l'EPA pour cette même année est de 12 %. Pour plus de détails concernant la méthodologie de l'EPA, voir la référence [5].

2.3 La non-réponse selon les strates de revenu

Il est impossible de comparer le taux de non-réponse selon le revenu car cette information n'est pas fournie par les non-répondants. Toutefois, le plan d'échantillonnage de l'EPA, utilisé pour l'EDA, a été conçu de sorte à former, dans neuf grandes villes, des strates d'aires géographiques où le revenu moyen des ménages excède 100 000 $ ainsi que dans sept grandes villes, des strates composées d'appartements habités par des ménages ayant un revenu moyen inférieur à 20 000 $. Le nombre de telles strates est petit et compte pour seulement un petit nombre de logements dans l'échantillon de l'EDA (environ 530 et 75 pour les strates de revenus élevés et faibles respectivement, soit un peu plus de 6 % de l'échantillon).

Tableau 2.3
Comparaison des taux de non-réponse (%) et taux de vacance (%) des strates à revenu élevé et à faible revenu par rapport aux autres strates, par trimestre et pour l'année, Canada (10 provinces)

Type de strate en fonction du revenu		TOTAL	Revenu élevé	Régulières	Revenu faible
Janvier à décembre	Vacants	7,5	3,2	7,8	6,7
	Non-réponse	28,5	37,2	27,8	30,4
	Pas de contact	11,4	16,0	11,1	13,0
	Refus	17,0	21,2	16,7	17,4
Janvier à mars	Vacants	8,2	5,6	8,4	0,0
	Non-réponse	34,7	44,3	34,0	28,6
	Pas de contact	15,0	24,2	14,4	0,0
	Refus	19,7	20,1	19,6	28,6
Avril à juin	Vacants	7,7	0,8	8,2	5,3
	Non-réponse	29,7	41,0	28,8	38,9
	Pas de contact	11,3	18,0	10,8	5,6
	Refus	18,5	23,0	18,0	33,3
Juillet à septembre	Vacants	7,3	3,3	7,4	15,8
	Non-réponse	27,0	35,7	26,4	25,0
	Pas de contact	10,5	13,4	10,2	25,0
	Refus	16,5	22,3	16,3	0,0
Octobre à décembre	Vacants	6,9	2,5	7,2	4,3
	Non-réponse	21,3	25,6	20,9	28,6
	Pas de contact	8,3	6,0	8,4	19,0
	Refus	13,0	19,7	12,6	9,5

Lorsqu'on examine pour l'année, le taux de non-réponse le plus élevé se retrouve dans les strates à revenu élevé (37,2 %) suivi des strates à faible revenu (30,4 %) et des strates régulières (27,8 %). Ce même ordre est respecté pour les non-contacts et pour les refus. La première cause de non-réponse est les refus.

Toujours pour l'année, le taux de vacance le plus élevé se retrouve dans les strates régulières (7,8 %).

Au niveau trimestriel, les taux de non-réponse se comportent différemment d'un trimestre à l'autre. Tout particulièrement les taux de refus des deux premiers trimestres sont les plus élevés dans les strates à faible revenu.

2.4 L'ajustement pour la non-réponse

Pour compenser la non-réponse, les poids de l'EDA sont gonflés par l'inverse du taux de réponse pondéré à l'intérieur de certains groupes définis dans chaque région par trimestre. Parmi les groupes définis dans chaque région, on y retrouve principalement les strates à revenu élevé, certaines régions métropolitaines et les différents niveaux d'urbanisation. Les taux pondérés diffèrent des taux présentés dans cette section puisqu'ils tiennent compte du poids de sondage de chaque ménage. Une description algébrique de l'ajustement pour la non-réponse est présentée à l'annexe A.

L'ajustement des poids pour la non-réponse permet de tenir compte des différences au niveau de la non-réponse par groupe défini. Il permettra de réduire le biais dans la mesure où les caractéristiques des répondants et des non-répondants sont similaires pour un même groupe.

3. Erreurs de couverture

Lors de la conception de l'enquête, on a défini la population visée, qu'on appelle la population cible. Rappelons d'abord la définition de cette population cible pour l'EDA puisqu'une bonne compréhension de la population cible est nécessaire à une bonne interprétation des données de l'enquête. Il est important de préciser que l'EDA utilise la base de sondage de l'Enquête sur la population active (EPA).

La population cible

> La population cible correspond aux individus des dix provinces et des villes de Whitehorse, Yellowknife et Iqaluit vivant dans les ménages privés au Canada. Les pensionnaires d'établissements institutionnels tels que les prisons, les hôpitaux pour malades chroniques, les résidences pour personnes âgées ainsi que les membres d'ordres religieux et d'autres groupes vivant en communauté, les membres des Forces Armées vivant dans les camps militaires et les individus vivant de façon permanente dans les hôtels ou les maisons de chambres sont donc exclus. On exclut aussi les représentants officiels de pays étrangers qui vivent au Canada et leurs familles ainsi que les résidents des réserves indiennes et des terres publiques.

Nous n'avons pas recueilli de renseignements auprès des personnes qui vivent temporairement loin de leur famille (par exemple, les étudiants d'université) parce que nous obtenons l'information auprès de leur famille, si elle est choisie dans l'échantillon.

De plus en 2001, pour des raisons pratiques principalement fondées sur le fait que la taille de l'échantillon était beaucoup plus petite que lors des enquêtes précédentes, les ménages habitant des logements situés dans les régions éloignées ont été exclus de l'enquête. Ces régions, en grande partie peu peuplées, sont situées principalement dans la portion nordique de certaines provinces. Elles sont difficilement accessibles et la collecte de données auprès de ces ménages est très dispendieuse.

L'enquête couvre donc près de 97 % de la population dans les dix provinces. Au Yukon, Territoires du Nord-Ouest et Nunavut, l'enquête ne couvre que les villes de Whitehorse, Yellowknife et Iqaluit ce qui représente 62 %, 44 % et 17 % de la population de leur territoire respectif selon le Recensement de 1996.

Les erreurs de couverture proviennent d'une représentation inadéquate de la population cible à partir des unités de la base de sondage. Certaines unités de la population cible peuvent être omises de la base de sondage, on parle alors d'un sous-dénombrement. D'autres unités qui ne sont pas dans la population cible peuvent être incluses par erreur ou certaines unités peuvent être incluses plus d'une fois, ces unités sont responsables du surdénombrement.

3.1 Le sous-dénombrement et le surdénombrement : les taux de glissement

Dans l'EDA, la sélection de l'échantillon est effectuée en utilisant une liste de logements dans chaque grappe sélectionnée. L'omission de logements lors de la création de la liste, les nouveaux logements qui s'ajoutent entre la création de la liste et la visite des intervieweurs (principalement dans les secteurs en développement) ainsi que la classification erronée de logements vacants contribuent au sous-dénombrement. Le fait d'inclure des logements qui ne seraient pas dans les limites de la grappe est une source de surdénombrement. De façon similaire, des erreurs peuvent se glisser lors de la collecte des données, lors de l'identification des personnes qui sont membres du ménage sélectionné. Ces erreurs contribuent également au sous-dénombrement et au surdénombrement.

Une bonne représentation de la population cible est indispensable à la production d'estimations de dépenses réalistes. Le nombre de personnes par ménage est également une caractéristique importante dans l'estimation des dépenses moyennes des ménages. Il est donc nécessaire que l'échantillon, non seulement représente adéquatement les individus de la population cible, mais également la distribution des ménages selon leur taille.

On observe généralement un sous-dénombrement net du nombre de personnes et du nombre de ménages dans l'EDA que l'on rectifie par un ajustement des poids en utilisant des données auxiliaires, qui sont basées sur des estimations démographiques post-censitaires. Le taux de glissement (voir la description algébrique en annexe A) est une mesure du pourcentage d'écart entre les estimations provenant de ces données auxiliaires et les estimations de l'enquête calculées en utilisant des poids non ajustés

avec ces données[4]. Les taux de glissement par groupe d'âge à l'échelle nationale et régionale sont présentés dans le tableau 3.1 alors que les taux de glissement selon les tailles de ménages se trouvent dans le tableau 3.2. Un taux positif correspond à un surdénombrement du nombre de personnes ou de ménages dans l'enquête.

Tableau 3.1
Taux de glissement (%) selon le groupe d'âge, cinq grandes régions canadiennes et Canada (10 provinces)

Âge	Canada (10 provinces)	Atlantique	Québec	Ontario	Prairies	Colombie-Britannique
0-6 ans	4,7	5,5	8,6	3,0	6,1	0,9
7-17 ans	1,5	2,2	-3,9	2,7	0,9	7,4
18-34 ans	-13,9	-19,8	-10,7	-17,9	-5,7	-15,3
35-54 ans	-7,7	-7,7	-11,0	-7,2	-4,9	-6,4
55-64 ans	-6,5	-19,7	1,0	-7,3	-6,4	-11,1
65 ans et +	-8,9	-6,6	-6,4	-5,8	-16,1	-15,0
Total	-6,9	-9,2	-6,6	-7,3	-4,6	-7,6

Pour l'EDA de 2001, le taux de sous-dénombrement de la population est de 6,9 %. En analysant le tableau 3.1 au niveau des groupes d'âge, on constate qu'à l'échelle nationale et régionale, les taux de glissement pour les enfants (0 à 6 ans et 7 à 17 ans) sont très différents de ceux des autres groupes d'âge. En effet, chez les enfants, on observe un surdénombrement ou un léger sous-dénombrement, alors que chez les adultes il y a toujours sous-dénombrement à l'exception du Québec pour le groupe des 55 à 64 ans.

Le taux de sous-dénombrement le plus élevé à l'échelle nationale est observé chez le groupe des 18 à 34 ans (13,9 %). À l'échelle régionale à l'exception des Prairies, ce groupe a un taux de sous-dénombrement élevé. On retrouve également un taux de sous-dénombrement élevé dans le groupe des 35 à 54 ans au Québec (11,0 %), dans le groupe des 55 à 64 ans de la région de l'Atlantique (19,7 %) et dans le groupe des 65 ans et plus dans les Prairies et en Colombie-Britannique (16,1 % et 15,0 % respectivement).

4. On utilise les sous-poids, c'est-à-dire les poids de sondage ajustés pour la non-réponse (voir Annexe A).

Tableau 3.2
Taux de glissement (%) selon la taille de ménage, cinq grandes régions canadiennes et Canada (10 provinces)

Région géographique	Ménages	Ménages d'une personne	Ménages de deux personnes	Ménages de trois personnes et plus
Canada (10 provinces)	-7,9	-16,9	-3,3	-5,8
Atlantique	-8,2	-13,8	2,9	-14,0
Québec	-7,4	-17,2	0,4	-6,1
Ontario	-8,5	-16,5	-5,1	-6,5
Prairies	-5,0	-14,1	-2,4	-1,6
Colombie-Britannique	-10,4	-22,2	-9,9	-3,1

On observe une sous-estimation du nombre de ménages de 7,9 % à l'échelle nationale. Ce taux de glissement est 1 point de pourcentage plus élevé que celui au niveau des personnes qui est de 6,9 %, voir le tableau 3.1.

À l'échelle régionale, le taux de sous-dénombrement varie de 5,0 % à 10,4 % pour le nombre de ménages. À l'exception de l'Atlantique, le taux de sous-dénombrement est beaucoup plus élevé chez les ménages d'une personne que pour les ménages de deux et de trois personnes ou plus. Dans le cas de l'Atlantique, les ménages d'une personne et ceux de trois personnes ou plus ont des taux de sous-dénombrement semblables, soit 13,8 % et 14,0 % respectivement.

3.2 L'ajustement au niveau de la population et des ménages

Afin de corriger le problème de représentativité illustré dans le tableau 3.1 et de réduire le biais qui en découle, on ajuste les données de l'enquête lors de la pondération en utilisant les estimations démographiques pour les groupes d'âge définis dans ce tableau, et ce pour chaque grande région canadienne. Cet ajustement réduit le biais mais ne l'élimine pas complètement si les caractéristiques des individus omis ou non-répondants diffèrent de celles des répondants inclus pour un même groupe d'âge dans une région.

On notera également que l'efficacité de l'ajustement à partir des estimations démographiques dépend beaucoup de la qualité de ces estimations démographiques et de leur exactitude à bien représenter la population cible de l'enquête. Les estimations démographiques ne sont pas exemptes d'erreurs. Ce sont des estimations post-censitaires basées sur les comptes de population du recensement de 1996, ajustées pour le sous-dénombrement net et qui tiennent compte des récentes statistiques sur la migration, les naissances, la mortalité, etc. Ces estimations démographiques sont ajustées pour tenir compte de certaines exclusions spécifiques aux enquêtes auprès des ménages comme les personnes vivant en institutions. Conceptuellement, elles diffèrent légèrement de la population cible de l'EDA en incluant les personnes habitant les ménages collectifs qui ne sont pas des institutions, par exemple les membres de groupes vivant en communauté et les individus vivant de façon permanente dans les hôtels ou les maisons de chambres. Cependant, on considère cette différence

négligeable puisque ces individus représentent moins de 0,4 % de la population canadienne.

Pour corriger le problème de représentativité de l'échantillon par rapport au nombre de ménages selon leur taille, illustré dans le tableau 3.2, on ajuste les données de l'enquête à l'aide de données auxiliaires. En ajustant les poids de l'EDA pour refléter les estimations post-censitaires du nombre de ménages selon la taille pour chaque grande région canadienne, on vise à compenser le biais qui découle d'une représentation inadéquate des ménages. Toutefois on n'éliminera pas nécessairement ce biais si les caractéristiques des ménages non interviewés (omis ou non-répondants) diffèrent de celles des ménages répondants pour une même taille de ménage. Comme dans le cas des estimations démographiques de population, l'efficacité de l'ajustement dépend de la qualité des données auxiliaires sur le nombre de ménages.

En plus des estimations démographiques citées précédemment, deux autres ensembles de données auxiliaires sont utilisés pour ajuster les données d'enquêtes lors de la pondération en vue d'en améliorer la représentativité. Le premier ensemble de données est utilisé pour contrôler le nombre d'enfants et d'adultes dans certaines grandes villes. Le second est conçu pour contrôler le nombre de ménages monoparentaux et de couples avec enfants pour chaque grande région canadienne.

Pour plus de détails sur la méthodologie de l'ajustement incluant celui des villes du Nord, voir la référence [1].

4. Les erreurs de réponse

Les erreurs de réponse correspondent au manque d'exactitude des réponses aux questions. Elles peuvent être attribuables à différents facteurs parmi lesquels se retrouvent un questionnaire qui nécessite des améliorations, une mauvaise interprétation des questions par les intervieweurs ou les répondants ainsi que des erreurs dans les déclarations des répondants.

Les répondants doivent inscrire leurs dépenses au fur et à mesure dans les deux journaux hebdomadaires qui leurs sont remis. Toutefois, des erreurs peuvent survenir, lorsque les répondants font des erreurs de transcription, n'utilisent pas les bons termes pour les produits ou devinent les montants s'il n'y a pas d'étiquette de prix ou de quantité d'un produit.

Plusieurs caractéristiques de l'enquête visent à aider les répondants à donner les renseignements les plus exacts possible. À l'occasion des visites personnelles, l'intervieweur donne des indications sur la façon de bien remplir le journal. Parfois, il aide le répondant à inscrire les données qui manquent. En outre, pour 2001, les répondants ont été fortement encouragés à joindre leurs reçus d'achat à la page du journal correspondant au jour de l'achat. Il y a eu aussi un autre élément nouveau en 2001, soit le carnet. Tous les membres du ménage âgés de 10 ans et plus disposaient d'un petit carnet à apporter constamment avec soi pour y inscrire les produits au moment de l'achat.
On considère également que le fardeau imposé au répondant, qui est principalement de rapporter une grande variété d'articles sur une période de deux semaines, peut

entraîner la fatigue du répondant et avoir un impact sur la qualité des réponses obtenues. En raison de ce fardeau associé au journal (FE3), on pose l'hypothèse que l'estimation des aliments achetés au magasin selon le questionnaire est plus fiable que l'estimation correspondante du journal.

En 2001, l'estimation des dépenses obtenue par l'entremise du questionnaire (FE2) est de 8,3 % supérieure à celle obtenue avec les données des journaux (FE3). Suivant l'hypothèse mentionnée ci-dessus, les valeurs pour les aliments achetés au magasin localement et à l'occasion de déplacements d'une seule journée dans chaque ménage ont été multipliées par un facteur de rajustement de 1.083.

Quoique l'on considère que les erreurs de réponses sont une source importante d'erreur, elles demeurent l'aspect lié à la qualité des données le plus difficile à mesurer. Il faut généralement entreprendre des études spéciales très coûteuses pour tenter de les mesurer. On tente de les réduire en implantant différentes procédures au cours du traitement telles que la vérification par exemple.

5. Les erreurs de traitement

Des erreurs de traitement peuvent se produire à toutes les étapes du traitement des données. Les principales étapes sont le codage des réponses, la saisie des données, la vérification, l'imputation pour la non-réponse partielle et la pondération. Dans l'EDA, on applique des procédures différentes à chaque étape, afin de réduire au minimum les erreurs de traitement, et on compare les estimations de l'enquête avec d'autres sources de données avant la diffusion. Les erreurs liées aux rajustements à l'étape de la pondération ont été décrites dans les sections 2 et 3. Les autres types d'erreurs de traitement sont abordés dans la présente section.

Le codage constitue une étape importante de l'EDA. Tous les articles achetés qui sont décrits doivent être codés. Le codage est effectué par des codeurs au bureau central de Statistique Canada, au moyen d'un logiciel élaboré à la Division de la statistique du revenu, qui les aide à choisir les codes appropriés d'aliments. Au cours du codage, chaque question et journal fait l'objet d'une série de vérifications et d'imputations manuelles. Certaines vérifications automatisées sont aussi effectuées au cours du codage, par exemple, pour déterminer les incohérences entre les codes et les unités de mesure. Par ailleurs, certains questionnaires sélectionnés de façon aléatoire sont vérifiés par des codeurs principaux.

Du point de vue de l'entrée des données, la charge de travail d'un préposé à la saisie est constituée de vingt questionnaires répartis en quatre lots de cinq. Un lot est sélectionné de façon aléatoire. Si la qualité des données du lot est acceptable, l'ensemble des questionnaires est approuvé; autrement, les trois autres lots sont vérifiés.

Comme il a été mentionné précédemment, au cours du codage, les questionnaires et journaux font l'objet de vérifications manuelles et automatisées, ainsi que d'une imputation manuelle. Un processus de vérification automatisée suit les étapes manuelles auxquelles sont soumis chaque questionnaire et chaque journal.

Ce processus de vérification automatisée permet de déterminer certaines données qui doivent être imputées, pour la plupart des données manquantes. Parmi les exemples de données manquantes dans le questionnaire (FE2) figurent les aliments achetés dans un restaurant hors du lieu de résidence – pendant une absence d'une nuit ou plus au cours du mois précédent, sans indication du nombre de repas achetés. Parmi les exemples de données manquantes dans le journal (FE3) figure la déclaration d'un achat de framboises, sans indication du type de conditionnement (fraîches, congelées, etc.). À cette étape, l'imputation est automatisée. On a recours à plus d'une méthode d'imputation, mais la méthode de base utilisée en 2001 est celle du plus proche voisin.

Cette technique repose sur la constitution de groupes de ménages similaires, selon certains critères (p. ex., province de résidence, trimestre de la collecte des données). À l'intérieur de ces groupes, chaque ménage qui doit être imputé (receveur) est apparié à un ménage qui a rempli un questionnaire (donneur) et qui s'apparente le plus à l'autre du point de vue de certaines caractéristiques (p. ex., revenu, taille, etc.). Les données du donneur sont imputées au receveur, à condition qu'elles répondent aux exigences d'uniformité avec les données déclarées par le receveur. Par exemple, si le receveur a indiqué avoir acheté des aliments préparés, les dépenses imputées pour cette variable doivent être supérieures à zéro.

Si on ne peut trouver de donneurs pour certains receveurs lors de cette première étape, on a recours encore une fois à la méthode d'imputation par donneur, en supprimant l'exigence liée à la similitude entre les revenus des ménages, qui est considérée comme moins importante que les autres caractéristiques. Plusieurs étapes d'imputation par donneur sont nécessaires afin de pouvoir imputer tous les receveurs. Dans quelques cas, la condition la plus importante, à savoir le trimestre, est aussi abandonnée pour trouver un donneur pour les cas qui restent. Pour plus de détails concernant les différentes étapes d'imputation par donneur, voir la référence [6].

Dans le cas du questionnaire (FE2), deux sections sont imputées : *Habitudes de dépense du ménage* (section B) et *Aliments et boissons consommés hors du lieu de résidence – pendant une absence d'une nuit ou plus au cours du mois précédent* (section C). Ces sections sont imputées séparément, étant donné qu'elles sont considérées comme relativement indépendantes l'une de l'autre. L'approche du plus proche voisin est utilisée dans la section B, mais pour la section C, il est parfois nécessaire d'utiliser d'autres méthodes, en raison de la difficulté de trouver un donneur approprié. Par exemple, certains répondants ont déclaré des dépenses au restaurant, mais n'ont pas déclaré le nombre de repas achetés, ou vice versa. Ces cas sont imputés grâce à la méthode d'imputation par la moyenne. La moyenne utilisée pour imputer est celle au niveau national.

Dans le cas du journal (FE3), pour chaque jour, on impute uniquement la section des *Aliments et boissons achetés dans un magasin*. Afin d'imputer cette section, on procède à des imputations indépendantes selon le regroupement des données manquantes qui doivent être imputées. De façon générale, l'imputation est effectuée au niveau de l'entrée, mais dans certains cas, elle se fait au niveau de la journée. Ces cas surviennent lorsque le répondant a indiqué uniquement les dépenses totales au titre des aliments achetés dans un magasin pour une journée donnée.
Pour que l'imputation au niveau de l'entrée soit possible, le répondant doit avoir fourni des renseignements additionnels. Par exemple, le répondant indique avoir acheté du bœuf, sans préciser le type : haché, en cubes, etc. Dans ce cas, le prix à l'unité est

considéré comme le critère d'appariement le plus important pour trouver un donneur. Si le prix à l'unité est manquant, on utilise la dépense totale de même que d'autres variables pour apparier.

Certaines entrées déclarées sont imprécises : *Légumes*, par exemple. Pour la première fois, en 2001, ces entrées sont codées et imputées en utilisant la méthode du plus proche voisin. Quelques fois, on ne peut pas trouver de donneurs pour certains receveurs, du fait d'une dépense élevée correspondant à l'entrée. Pour l'imputation, on additionne pour chaque donneur possible toutes les dépenses d'une journée correspondant aux entrées appropriées (tous les légumes par exemple). Lorsqu'un donneur est trouvé avec ses dépenses additionnées correspondant aux dépenses du receveur, on attribue toutes les entrées appropriées d'aliments au receveur.

Lorsque le répondant a indiqué uniquement les dépenses totales au titre des aliments achetés dans un magasin pour une journée donnée, l'imputation est effectuée au niveau de la journée. On apparie les dépenses d'une journée du receveur à celles du donneur. Les journées qui peuvent servir de donneur sont celles qui ne comportent pas d'imputations au niveau de l'entrée. Toutes les entrées pour une journée d'un donneur sélectionné sont attribuées au receveur. Les dépenses totales du receveur sont remplacées par les dépenses totales du donneur dans les cas où les deux chiffres ne correspondent pas exactement.

Enfin, il convient de souligner que les ménages qui n'ont fourni qu'une semaine de données ne font pas l'objet d'une imputation dans le cas de la semaine pour laquelle ils n'ont pas déclaré de données.

Le biais découlant de l'imputation pour la non-réponse partielle est difficile à évaluer. Il dépend des différences entre les répondants et les non-répondants, ainsi que de la capacité de la méthode d'imputation de produire des estimations non biaisées. Toutefois, les proportions de ménages ayant des données imputées et les taux d'imputation montrent l'importance de la non-réponse partielle. L'incidence de l'imputation est une bonne indication du biais potentiel des estimations. Ces indicateurs de qualité sont présentés à la section suivante.

5.1 Indicateurs de qualité provenant de l'imputation

La proportion des ménages qui doivent être imputés, et le nombre de variables imputées par ménage, constituent une indication préliminaire de l'ampleur de la non-réponse partielle. Parmi les autres indications figurent les taux d'imputation et l'incidence de l'imputation. L'incidence est définie comme étant une mesure de l'impact des valeurs imputées sur les estimations. Une description algébrique du calcul de l'incidence de l'imputation est présentée à l'annexe A.

Les réponses des ménages peuvent être réparties en deux catégories : celles recueillies à partir du questionnaire (FE2) et celles recueillies à partir du journal (FE3). Les différents indicateurs de qualité dans le cas du questionnaire (FE2) et du journal (FE3) sont présentés dans les deux prochaines sous-sections, respectivement.

5.1.1 Questionnaire (FE2)

Seulement deux sections du FE2 sont imputées : *Habitudes de dépense du ménage* (section B) et *Aliments et boissons consommés hors du lieu de résidence – pendant une absence d'une nuit ou plus au cours du mois précédent* (section C). On peut imputer jusqu'à cinq variables dans la section B (questions 1 à 4 et 4.1), et jusqu'à 50 dans la section C (questions 1 à 6).

Le pourcentage de ménages répondants[5] qui nécessitent une imputation dans le cas de la section B est présenté au tableau 5.1, aux niveaux régional et national. Le tableau est ventilé selon le nombre de variables imputées (sur 5) pour un ménage.

Tableau 5.1
Pourcentage de ménages imputés pour la section B du questionnaire (FE2), cinq grandes régions canadiennes et Canada (10 provinces)

Région géographique	Nombre de variables imputées			Total
	1	2	3 à 5	
Canada (10 provinces)	0,8	10,8	0,6	12,1
Atlantique	0,4	8,4	0,0	8,8
Québec	0,5	3,6	0,0	4,1
Ontario	0,6	15,4	0,7	16,6
Prairies	1,2	13,0	1,2	15,5
Colombie-Britannique	1,1	11,4	0,8	13,3

Le tableau 5.1 indique que près de 88 % des ménages au niveau du Canada n'ont pas besoin d'imputation pour la section B. La plupart des ménages pour lesquels une imputation est nécessaire comportent deux variables à imputer, habituellement les questions 4 et 4.1 portant sur les dépenses au titre des aliments préparés pour des occasions spéciales. Les régions de l'Ontario (16,6 %), des Prairies (15,5 %) et de la Colombie-Britannique (13,3 %) ont les pourcentages les plus élevés d'imputation, tandis qu'au Québec le pourcentage est le plus faible (4,1 %).

Les incidences de l'imputation calculées pour les quatre variables quantitatives de la section B sont présentées au tableau 5.2, aux niveaux régional et national. L'incidence de l'imputation est définie comme le total des valeurs imputées pondérées, divisé par le total des valeurs pondérées (comprenant les valeurs imputées et non imputées).

5. Les ménages répondants correspondent à tous les ménages vivant dans des logements admissibles, à l'exclusion des ménages qui n'ont pu être contactés ou qui ont refusé de participer à l'enquête.

Tableau 5.2
Incidences de l'imputation (%) pour la section B du questionnaire (FE2), cinq grandes régions canadiennes et Canada (10 provinces)

Région géographique	Aliments et autres produits alimentaires (Question 1)	Articles non alimentaires (Question 2)	Aliments en vrac (Question 3)	Aliments préparés pour des occasions (Question 4.1)
Canada (10 provinces)	0,8	1,0	0,0	12,2
Atlantique	0,4	0,7	0,0	23,9
Québec	0,2	0,3	0,0	1,2
Ontario	0,7	0,7	0,0	15,5
Prairies	1,6	2,5	0,0	14,5
Colombie-Britannique	1,8	0,9	0,0	4,5

Le tableau 5.2 indique que la variable des « Dépenses au titre des aliments préparés pour des occasions » (question 4.1) est celle qui a l'incidence la plus élevée. Il s'agit de la variable de la section qui comporte le taux le plus élevé d'imputation ainsi que celle ayant peu de ménages déclarant ce type de dépenses. Au niveau du Canada, cette variable a une incidence de 12,2 %, tandis que les trois autres variables ont une incidence de moins de 1,0 %. Au niveau régional, si nous examinons uniquement l'incidence de l'imputation dans le cas des « Dépenses au titre des aliments préparés pour des occasions », c'est dans la région de l'Atlantique (23,9 %) que celle-ci est la plus élevée, suivie par l'Ontario (15,5 %) et les Prairies (14,5 %). C'est au Québec que l'incidence est la plus faible, à 1,2 %.

Le pourcentage de ménages répondants qui nécessitent une imputation à la section C est présenté au tableau 5.3, aux niveaux régional et national. Le tableau est ventilé selon le nombre de variables imputées (sur 50) pour un ménage.

Tableau 5.3
Pourcentage de ménages imputés pour la section C du questionnaire (FE2), cinq grandes régions canadiennes et Canada (10 provinces)

Région géographique	Nombre de variables imputées					Total
	1 à 10	11 à 20	21 à 30	31 à 40	41 à 50	
Canada (10 provinces)	1,6	0,0	2,1	0,0	0,4	4,2
Atlantique	1,3	0,0	1,8	0,0	0,0	3,1
Québec	0,9	0,0	1,3	0,0	0,2	2,3
Ontario	0,9	0,1	2,1	0,0	0,8	3,9
Prairies	2,9	0,0	2,6	0,0	0,4	5,8
Colombie-Britannique	2,0	0,0	3,1	0,1	0,3	5,6

Selon le tableau 5.3 au niveau du Canada, plus de 95 % des ménages n'ont pas besoin d'imputation pour la section C. Pour la moitié des ménages qui doivent être imputés, l'imputation touche de 21 à 30 variables (principalement la question 4 concernant les aliments et les boissons non alcoolisées achetés dans un restaurant hors du lieu de

résidence – pendant une absence d'une nuit ou plus, qui comprend 28 données à déclarer), et pour la majorité de l'autre moitié, l'imputation touche de 1 à 10 variables. Au niveau régional, près de 6 % des ménages des Prairies et de la Colombie-Britannique doivent être imputés. Le Québec a le pourcentage le plus faible, à 2,3 %.

Les incidences de l'imputation calculées pour les 31 variables quantitatives, à l'exclusion des questions 1 et 2[6] de la section C, sont présentées au tableau 5.4, aux niveaux régional et national. Étant donné ce nombre élevé de variables, plutôt que de présenter l'incidence de l'imputation pour chacune d'elles, on montre le nombre de variables par catégorie d'incidence (%).

Tableau 5.4
Incidences de l'imputation (%) pour la section C du questionnaire (FE2), cinq grandes régions canadiennes et Canada (10 provinces)

Région géographique	Nombre de variables selon la catégorie d'incidence (%) pour la section C du FE2					
	0 à 5 %	5 à 10 %	10 à 15 %	15 à 25 %	25 à 35 %	35 % ou plus
Canada (10 provinces)	8	15	5	2	1	0
Atlantique	13	8	5	2	1	2
Québec	22	4	2	3	0	0
Ontario	15	4	7	1	3	1
Prairies	12	8	3	3	2	3
Colombie-Britannique	12	3	6	8	2	0

Selon le tableau 5.4, au niveau du Canada, 23 variables sur 31 ont une incidence de moins de 10 %. Il n'y a que 3 variables qui ont des incidences supérieures à 15 % et elles se retrouvent dans la catégorie « Autre » de la question 4, à savoir : les *Dîners* (33,3 %), les *Soupers* (22,5 %) et les *Déjeuners* (21,7 %). Au niveau régional, la Colombie-Britannique est la région ayant le nombre le plus élevé de variables (10) avec des incidences de plus de 15 % suivie par la région des Prairies avec 8 variables. C'est au Québec que les incidences sont les plus faibles.

5.1.2 Journal (FE3)

Pour chaque jour, seule la section *Aliments et boissons achetés dans un magasin* est imputée dans le FE3.

Les 5 999 ménages répondants ont déclaré 279 709 dépenses, et 2,2 % d'entre elles sont imputées. La plupart du temps, l'imputation précise le type de dépense. Par exemple, si le répondant indique qu'il a acheté du lait, sans spécifier le type de lait, cette entrée sera imputée parmi les choix : lait entier de consommation, lait à faible teneur en gras (1 %), lait à faible teneur en gras (2 %), lait écrémé ou produits spéciaux du lait.

Le pourcentage de ménages répondants qui doivent être imputés en ce qui a trait au type de dépenses dans les journaux (FE3) est présenté dans le tableau 5.5, aux niveaux régional et national. Ces proportions sont élevées (38 % au niveau national) parce que

6. L'incidence de l'imputation n'a été calculée que pour les variables représentant des dépenses alimentaires ou des nombres de repas.

chaque ménage doit dresser une liste d'un nombre important d'entrées au cours de la période de déclaration de deux semaines. Tous les ménages qui ont une imputation dans l'une ou l'autre de ces entrées sont considérés comme des ménages imputés.

Tableau 5.5
Pourcentage de ménages imputés pour les journaux (FE3), cinq grandes régions canadiennes et Canada (10 provinces)

Région géographique	Nombre d'imputations requises			Total
	1	2	3 ou plus	
Canada (10 provinces)	13,1	8,3	16,6	38,0
Atlantique	14,0	6,0	9,5	29,5
Québec	16,1	10,8	18,8	45,6
Ontario	12,3	9,2	21,9	43,4
Prairies	11,6	6,5	12,3	30,5
Colombie-Britannique	11,9	7,4	14,8	34,2

Selon le tableau 5.5, au niveau national, des données sont imputées dans les journaux pour 38,0 % des ménages. Il est pertinent de noter, toutefois, que le tiers de ces ménages n'ont qu'une seule entrée d'imputée tandis que le nombre moyen d'entrées par ménage est de 49. Cette moyenne est basée seulement sur les ménages ayant déclarées des dépenses dans les journaux. Au niveau régional, le Québec (45,6 %) et l'Ontario (43,4 %) ont les pourcentages les plus élevés d'imputation, mais dans le cas des ménages avec trois imputations ou plus, l'Ontario a un pourcentage de ménages imputés supérieur au Québec. C'est dans l'Atlantique que le pourcentage est le plus faible, à 29,5 %.

On code le type d'aliment pour chacune des 279 709 dépenses déclarées. Ce codage pour l'EDA de 2001 a produit 205 codes différents. Les taux d'imputation et les incidences de l'imputation pour les 11 dépenses les plus fréquemment déclarées sont présentés respectivement dans les tableaux 5.6 et 5.7. Ces taux et incidences sont calculés par type de dépense. Les deux tableaux sont établis aux niveaux régional et national. Les tableaux sont dans l'ordre commençant par la dépense la plus fréquemment déclarée.

Tableau 5.6
Taux d'imputation (%) pour les journaux (FE3), cinq grandes régions canadiennes et Canada (10 provinces) pour les 11 dépenses les plus fréquemment déclarées

Description de l'aliment	Canada (10 prov)	Atlantique	Québec	Ontario	Prairies	Colombie-Britannique
Pain	3,0	0,5	2,1	4,2	4,3	2,4
Boissons gazeuses	1,6	0,1	1,4	2,1	3,2	0,1
Petits pains et brioches, non sucrés	2,8	1,0	2,1	3,4	4,3	1,4
Bananes et bananes des antilles - fraîches	3,7	0,7	3,5	5,2	4,4	2,4
Lait à faible teneur en gras (2 %)	15,6	15,6	13,8	19,9	14,7	11,7
Poulet (y compris poules) - frais ou congelé	5,3	0,7	4,2	8,5	4,4	4,6
Biscuits sucrés, secs, ordinaires, de fantaisie	4,0	0,6	2,2	7,5	4,9	2,1
Céréales de table (sauf céréales pour bébés)	2,1	0,4	1,2	2,4	4,4	1,3
Soupe - froide, congelée ou en boîte	2,2	0,3	2,2	2,5	3,6	1,0
Tomates - fraîches	6,5	0,0	5,6	11,3	5,8	2,9
Oeufs	2,1	0,4	2,2	2,2	3,5	1,4

Au niveau du Canada, le tableau 5.6 montre que la plupart de ces dépenses comportent ont un taux d'imputation inférieur à 5 %. Le taux d'imputation le plus élevé est observé pour le lait à faible teneur en gras (2 %), avec 15,6 %. L'imputation de ce type de dépense survient la plupart du temps avec des répondants qui ont déclaré une dépense pour le lait, mais n'ont pas précisé le type de lait.

Au niveau régional, les taux les plus élevés d'imputation pour ces diverses dépenses se retrouvent en Ontario et dans les Prairies. La région de l'Atlantique est celle où les taux d'imputation sont les plus faibles, à 1 % ou moins, si nous excluons les dépenses au titre du lait à faible teneur en gras (2 %).

Dans le tableau 5.6, le Québec est l'une des régions où les taux d'imputation sont faibles pour les 11 dépenses les plus fréquemment déclarées, même si cette province a le pourcentage de ménages nécessitant au moins une imputation le plus élevé pour les ménages, comme le montre le tableau 5.5. Cela peut s'expliquer par la répartition des données manquantes. Au Québec, on compte davantage de répondants pour lesquels il manque au moins une donnée dans le journal, mais les données manquantes sont réparties entre les différents types de dépenses de façon plus marquée au Québec comparativement aux autres régions.

Tableau 5.7
Incidences de l'imputation (%) sur la variable des dépenses des journaux (FE3), cinq grandes régions canadiennes et Canada (10 provinces) pour les 11 dépenses les plus fréquemment déclarées

Description de l'aliment	Canada (10 prov)	Atlantique	Québec	Ontario	Prairies	Colombie-Britannique
Pain	3,6	0,4	2,0	5,7	4,3	2,8
Boissons gazeuses	2,0	0,1	1,3	3,2	2,4	0,1
Petits pains et brioches, non sucrés	3,9	1,5	1,8	5,8	5,0	1,2
Bananes et bananes des antilles - fraîches	4,6	1,1	4,0	6,0	4,9	3,8
Lait à faible teneur en gras (2 %)	16,9	14,0	13,6	21,1	16,7	13,2
Poulet (y compris poules) - frais ou congelé	7,5	0,3	2,9	13,4	4,2	4,2
Biscuits sucrés, secs, ordinaires, de fantaisie	5,5	1,9	2,1	10,2	5,1	1,6
Céréales de table (sauf céréales pour bébés)	2,1	0,6	1,1	2,7	3,9	1,1
Soupe - froide, congelée ou en boîte	2,4	0,6	2,0	3,8	2,9	0,4
Tomates - fraîches	11,6	0,0	7,0	20,6	5,5	3,1
Oeufs	2,5	0,6	1,8	2,8	4,2	2,1

Selon le tableau 5.7, au niveau du Canada, pour sept de ces dépenses, l'incidence de l'imputation pour la variable des dépenses est inférieure à 5 %, deux ont des incidences se situant entre 5 % et 8 %, et deux ont des incidences se situant entre 11 % et 17 %. Au niveau régional, les incidences d'impact les plus grandes se retrouvent en grande partie en Ontario. Celles dans les régions de l'Atlantique, Québec et Colombie-Britannique sont inférieures au niveau national.

Pour des renseignements plus détaillés sur l'imputation, voir la référence [6].

Bibliographie

[1] Auger, S., Khan, K. et Tremblay J. (2004), *La méthodologie de l'Enquête sur les dépenses alimentaires*, Division des méthodes d'enquêtes auprès des ménages, N° 62F0026MIF-2005004 au catalogue, Statistique Canada.

[2] Wolter, K.M. (1985), *Introduction to Variance Estimation*, Springer-Verlag New-York Inc.

[3] *Dépenses alimentaires au Canada (2001)*, N° 62-554-XIF au catalogue.

[4] *Enquête sur les dépenses des ménages de 1997 – Indicateurs de qualité des données*, Division des méthodes d'enquêtes auprès des ménages, Document interne, Statistique Canada.

[5] *Méthodologie de l'enquête sur la population active du Canada*, N° 71-526-XPB au catalogue.

[6] Khan, K. (2002), *Imputation of the 2001 Food Expenditure Survey*, Division des méthodes d'enquêtes auprès des ménages, Document interne, Statistique Canada.

Annexe A

Notation algébrique

1. Ajustement pour la non-réponse

Le sous-poids (c-à-d. le poids de sondage ajusté pour la non-réponse) d'un ménage k, noté w_k^{NR}, est

$$w_k^{NR} = \pi_k^{-1} * \frac{1}{taux_g} \quad avec \quad taux_g = \frac{\sum_{k \in s_{g,r}} \pi_k^{-1}}{\sum_{k \in s_{g,r}} \pi_k^{-1} + \sum_{k \in s_{g,nr}} \pi_k^{-1}}$$

où

$s_{g,r}$ est l'ensemble des ménages répondants dans le groupe de non-réponse g,

$s_{g,nr}$ est l'ensemble des ménages non-répondants (refus, non-contacts) dans le groupe de non-réponse g, et

π_k^{-1} est le poids de sondage attribué au ménage k (π_k étant la fraction de sondage).

2. Calcul du taux de glissement

Le taux de glissement pour un groupe de contrôle c, noté $taux_c$, est

$$taux_c = 100 * \frac{\left(\sum_{k \in s_{c,r}} w_k^{NR}\right) - t_c}{t_c}$$

où

$s_{c,r}$ est l'ensemble des ménages répondants dans le groupe de contrôle c,

w_k^{NR} est le sous-poids du ménage k (voir la définition de w_k^{NR} ci-dessus), et

t_c est le total de la donnée auxiliaire pour le groupe de contrôle c.

3. Calcul de l'incidence de l'imputation

L'incidence de l'imputation pour un type de dépense a, noté rep_d_a, est

$$rep_d_a = \frac{\sum_{k \in s_{i,r}} w_k^f * d_{a,k}}{\sum_{k \in s_{i,r}} w_k^f * d_{a,k} + \sum_{k \in s_{ni,r}} w_k^f * d_{a,k}}$$

où

$s_{i,r}$ est l'ensemble des ménages répondants dont le type de dépense a est imputé,

$s_{ni,r}$ est l'ensemble des ménages répondants dont le type de dépense a n'est pas imputé,

w_k^f est le poids final attribué au ménage k., et

$d_{a,k}$ est le montant associé au type de dépense a du ménage k.

Annexe B
Questionnaire de l'Enquête sur les dépenses alimentaires de 2001 (FE2)

Enquête sur les dépenses alimentaires en 2001

Questionnaire

FE2

Confidentiel une fois rempli

Déclaration exigée en vertu de la Loi sur la statistique, Statuts revisés du Canada, 1985, chapitre S-19.

See reverse for English

Strate	Type	Grappe	Rés	Liste

M	Mois	N

Registre des livraisons et des reprises

Visite	Jour	Mois	Début	Fin
Interview et livraison des journaux			:	:
Reprise du premier journal			:	:
Reprise du deuxième journal			:	:

Registre des visites

Numéro de la visite	Date	Heure	Observations

Nom de l'intervieweur (En majuscules)

Numéro de l'intervieweur

Statistique Canada / Statistics Canada

Canada

Détachez et détruisez après le traitement

Adresse de liste

Contact du ménage

No de téléphone

Ind. rég.

A. Composition du ménage

	Personne 01	Personne
1 Quels sont les **prénoms** de tous les membres de votre ménage ? *Inscrivez d'abord la personne de référence du ménage (voir la définition). Posez toutes les questions de la section A à chaque membre du ménage que vous avez inscrit.* **Personne de référence du ménage** – le membre du ménage qui en est le principal soutien financier (c.-à-d. la personne qui paie le loyer, les versements hypothécaires, les taxes foncières ou l'électricité, etc.). Il peut s'agir d'un homme ou d'une femme. Si tous les membres se partagent les responsabilités financières à parts égales, inscrivez n'importe quel membre comme personne de référence du ménage.	Prénom	Prénom
2 Quel est le **lien de parenté** de _____ avec la personne de référence ?	(X) Personne de référence	( ) Conjoint(e) ( ) Fils ou fille ( ) Autre parenté ( ) Non apparenté(e)
3 En quels mois et année _____ est-il (elle) né(e) ? *Si appelé en 1960 ou avant, inscrivez 1900.*	M M A A A A	M M A A A A
4 Est-ce que _____ est un homme ou une femme ?	( ) Homme ( ) Femme	( ) Homme ( ) Femme
5 Quel est l'**état matrimonial de** _____ ? *Cocher un cercle.*	( ) Marié(e) à un membre du ménage ( ) Conjoint(e) de fait d'un membre du ménage ( ) Jamais marié(e) (célibataire) ( ) Autre (séparé(e), divorcé(e), veuf ou veuve)	( ) Marié(e) à un membre du ménage ( ) Conjoint(e) de fait d'un membre du ménage ( ) Jamais marié(e) (célibataire) ( ) Autre (séparé(e), divorcé(e), veuf ou veuve)
6 *Code de famille économique* **Famille économique** – deux personnes ou plus qui vivent dans le même logement et qui sont apparentées l'une à l'autre par le mariage (par adoption ou de vivent en union libre). Des cohabitantes non apparentées auront des code différents.	Inscrivez le code A	Inscrivez le code

Vérifiez la composition du ménage

Après avoir dressé la liste des membres du ménage, demandez

- Y a-t-il quelqu'un d'autre qui vit à cette adresse ?
- Y a-t-il des personnes absentes qui peuvent être considérées comme vivant à cette adresse ?

Exemplaire seulement

A. Composition du ménage

Personne	Personne	Personne	Personne
Prénom	Prénom	Prénom	Prénom
○ Conjoint(e) ○ Fils ou fille ○ Autre parent(e) ○ Non apparenté(e)	○ Conjoint(e) ○ Fils ou fille ○ Autre parent(e) ○ Non apparenté(e)	○ Conjoint(e) ○ Fils ou fille ○ Autre parent(e) ○ Non apparenté(e)	○ Conjoint(e) ○ Fils ou fille ○ Autre parent(e) ○ Non apparenté(e)
M M A A A A	M M A A A A	M M A A A A	M M A A A A
○ Homme ○ Femme	○ Homme ○ Femme	○ Homme ○ Femme	○ Homme ○ Femme
○ Marié(e) à un membre du ménage ○ Conjoint(e) de fait d'un membre du ménage ○ Jamais marié(e) (célibataire) ○ Autre (séparé(e), divorcé(e), veuf ou veuve)	○ Marié(e) à un membre du ménage ○ Conjoint(e) de fait d'un membre du ménage ○ Jamais marié(e) (célibataire) ○ Autre (séparé(e), divorcé(e), veuf ou veuve)	○ Marié(e) à un membre du ménage ○ Conjoint(e) de fait d'un membre du ménage ○ Jamais marié(e) (célibataire) ○ Autre (séparé(e), divorcé(e), veuf ou veuve)	○ Marié(e) à un membre du ménage ○ Conjoint(e) de fait d'un membre du ménage ○ Jamais marié(e) (célibataire) ○ Autre (séparé(e), divorcé(e), veuf ou veuve)
Inscrivez le code	*Inscrivez le code*	*Inscrivez le code*	*Inscrivez le code*

Notes et observations

Pour fins d'illustration seulement

B. Habitudes de dépense du ménage

En faisant abstraction des achats effectués lors d'une absence de plus d'une journée, au cours des quatre dernières semaines . . .	Coût total
1. Combien estimez-vous que le ménage a dépensé au titre des aliments et autres produits alimentaires achetés dans des magasins? (Comptez les aliments achetés à des comptoirs de produits agricoles ou livrés à domicile) • **Ne comptez pas** l'achat des aliments en vrac. *(Voir Q.3)*	$
2. De ce montant, combien le ménage a-t-il dépensé au titre des **articles non alimentaires** comme les produits de papier, les articles ménagers, la nourriture pour animaux domestiques, les boissons alcoolisées, etc. ?	$
3. Quel montant le ménage a-t-il dépensé au titre des achats d'aliments en vrac, par exemple, les quantités de viande DÉPASSANT 25 kg (55 lbs.), fruits ou légumes en vrac pour la mise en conserves, la congélation, etc. ? • **Comptez** les frais de découpage, d'emballage et de congélation.	$
4. Le ménage a-t-il acheté des aliments préparés ou des boissons non alcoolisées dans des magasins pour des réceptions, noces et autres occasions, qui ne sont pas indiqués à la Q.1 ou à la Q.3 ci-dessus ? • **Ne comptez pas** les restaurants et les traiteurs. 4.1 Quel montant a-t-il dépensé ?	○ Oui → *Continuez* ○ Non → *Passez à la Q. 5* $

5. Pour les **12 derniers mois**, ou se situait dans l'échelle suivante le revenu **total** provenant de toutes sources, de **tous** les membres de votre ménage, **avant** impôt et autres déductions ?	○ Moins que 5 000 $ ○ 5 000 $ - 9 999 $ ○ 10 000 $ - 14 999 $ ○ 15 000 $ - 19 999 $ ○ 20 000 $ - 29 999 $ ○ 30 000 $ - 39 999 $ ○ 40 000 $ - 49 999 $ ○ 50 000 $ - 59 999 $ ○ 60 000 $ - 69 999 $ ○ 70 000 $ - 79 999 $ ○ 80 000 $ - 99 999 $ ○ 100 000 $ ou plus ○ Ne sait pas ○ Refus

Notes et observations

Pour information seulement

C. Aliments et boissons consommés hors du lieu de résidence (pendant une absence d'une nuit ou plus au cours du mois précédent)

1. Des membres du ménage se sont-ils absentés du foyer une nuit ou plus au cours du mois de _____ ?
   (Intervieweur : indiquez le mois précédant ci-dessus)
   ○ Oui → *Continuez*
   ○ Non → *Fin de l'entrevue*

2. Pour chaque absence d'une nuit ou plus au cours du mois précédent, indiquez le nombre de personnes en cause et le nombre de nuits passées hors du foyer.

	Absence N° 1	Absence N° 2	Absence N° 3	Absence N° 4	Absence N° 5
Nombre de personnes absentes	002	004	006	008	010
Nuits passées hors du foyer	003	005	007	009	011

3. Au cours de ce séjour, quel est le montant de la pension que les membres du ménage ont payé à des ménages privés ? 012 $

4. Pour toutes ces absences, j'aimerais maintenant vous poser quelques questions sur les aliments et les boissons non alcoolisées ACHETÉS DANS UN RESTAURANT.
   - **Comptez** les pourboires, les taxes et les achats faits pour les invités.
   - **Ne comptez pas** les dépenses pour lesquelles vous avez été remboursé ou qui étaient comprises dans un forfait.

		Restaurants avec service aux tables	Restaurants à service rapide	Cafétérias	Autre
Déjeuners	Nombre de repas	013	014	015	016
	Dépenses	017 $	018 $	019 $	020 $
Dîners	Nombre de repas	021	022	023	024
	Dépenses	025 $	026 $	027 $	028 $
Soupers	Nombre de repas	029	030	031	032
	Dépenses	033 $	034 $	035 $	036 $
Casse-croûtes, friandises, boissons non alcoolisées	Dépenses	037 $	038 $	039 $	040 $

Nota : « Autres » restaurants incluent les comptoirs de rafraîchissements, les casse-croûtes, les distributeurs automatiques, les cantines mobiles, les vendeurs de pommes de terre frites, les traiteurs, les vendeurs de café, etc.

5. Au cours du mois précédent, combien ce ménage a-t-il dépensé au titre des aliments et des boissons non alcoolisées ACHETÉS AU MAGASIN pendant une absence d'une nuit ou plus ? Par exemple casse-croûtes achetés dans un poste d'essence, aliments achetés dans une épicerie, etc. 041 $

6. Combien de repas ont été reçus gratuitement, remboursés, ou encore étaient compris dans un forfait (p. ex., voyages d'affaires, avec des amis ou en famille) ? 042

D. Compte rendu du suivi

A compléter par l'intervieweur à chaque reprise de journal.	Première semaine	Deuxième semaine
1. a) Pendant ou après le suivi, avez-vous, ou le répondant a-t-il, inscrit des ALIMENTS ACHETÉS AU MAGASIN dans le journal ?	001 1 ○ Oui, tous les achats → ***Passez à la Q. 2*** 2 ○ Oui, **certains** achats *Veuillez estimer la valeur monétaire de ces items.* $ 002 → ***Passez à la Q. 2*** 3 ○ Non → ***Continuez***	009 1 ○ Oui, tous les achats → ***Passez à la Q. 2*** 2 ○ Oui, **certains** achats *Veuillez estimer la valeur monétaire de ces items.* $ 010 → ***Passez à la Q. 2*** 3 ○ Non → ***Continuez***
b) Si aucun achat n'a été ajouté, quelle en était la raison ?	003 1 ○ Le répondant avait déjà bien noté tous les achats 2 ○ Le répondant n'avait fait aucun achat au magasin 3 ○ Autre *Veuillez inscrire la raison à la Q. 5*	011 1 ○ Le répondant avait déjà bien noté tous les achats 2 ○ Le répondant n'avait fait aucun achat au magasin 3 ○ Autre *Veuillez inscrire la raison à la Q. 5*
2. a) Pendant ou après le suivi, avez-vous, ou le répondant a-t-il, inscrit des ALIMENTS ACHETÉS AU RESTAURANT dans le journal ?	004 1 ○ Oui, tous les achats → ***Passez à la Q. 3*** 2 ○ Oui, **certains** achats *Veuillez estimer la valeur monétaire de ces items.* $ 005 → ***Passez à la Q. 3*** 3 ○ Non → ***Continuez***	012 1 ○ Oui, tous les achats → ***Passez à la Q. 3*** 2 ○ Oui, **certains** achats *Veuillez estimer la valeur monétaire de ces items.* $ 013 → ***Passez à la Q. 3*** 3 ○ Non → ***Continuez***
b) Si aucun achat n'a été ajouté, quelle en était la raison ?	006 1 ○ Le répondant avait déjà bien noté tous les achats 2 ○ Le répondant n'avait fait aucun achat au restaurant 3 ○ Autre *Veuillez inscrire la raison à la Q. 5*	014 1 ○ Le répondant avait déjà bien noté tous les achats 2 ○ Le répondant n'avait fait aucun achat au restaurant 3 ○ Autre *Veuillez inscrire la raison à la Q. 5*
3. Quel est l'état final de la section du journal ALIMENTS ACHETÉS AU MAGASIN ?	007 1 ○ Complet 2 ○ Incomplet	015 1 ○ Complet 2 ○ Incomplet
4. Quel est l'état final de la section du journal ALIMENTS ACHETÉS AU RESTAURANT ?	008 3 ○ Complet 4 ○ Incomplet	016 3 ○ Complet 4 ○ Incomplet

Pour information seulement

5. Veuillez noter vos observations sur toutes les sections qui étaient incomplètes ou sur n'importe quel problème rencontré en remplissant le journal.

Première semaine
020

Deuxième semaine
021

Notes et observations

Pour information seulement

Annexe C
Journal de l'Enquête sur les dépenses alimentaires de 2001 (FE3) — Exemple d'une journée

Ouvrez les volets latéraux pour consulter les notes

Joindre les reçus ici

Jour 1

Cochez (✓) le jour approprié:

☐ LUN ☐ MAR ☐ MER ☐ JEU ☐ VEN ☐ SAM ☐ DIM

Aliments et boissons achetés dans un MAGASIN

Si vous n'avez rien acheté, cochez (✓) ici ☐

Décrivez l'article acheté • Donnez une description **complète** de chaque aliment acheté • **Évitez** les abréviations et écrivez clairement en lettres moulées • **Inscrivez** les achats en vrac	**Ce produit était-il . . . ?** A = frais B = congelé C = en boîte / embouteillé / en contenant de plastique D = séché E = autre **Inscrivez la lettre code**	**Acheté dans . . .** A = supermarché B = marché ou magasin d'alimentation spécialisé C = dépanneur D = autre **Inscrivez la lettre code**	**Nombre de . . .** conserves, bouteilles, cartons, sacs, boîtes, paquets, pommes, unités, douzaines, etc. **Inscrivez la quantité**	**Poids ou volume par unité** Exemples 1 L (1 litre) 2 L (2 litres) 500 G 1 KG 16 OZ UNITÉ DOUZAINE PAQUET	**Coût total** **Exclure** toutes les taxes $	c	**À l'usage du bureau**
Exemple: COLA	C	A	24	355 ML	6	99	

Ouvrez les volets latéraux pour consulter les notes

Jour 1 *suite*

Aliments et boissons achetés dans un MAGASIN

Décrivez l'article acheté • Donnez une description **complète** de chaque aliment acheté • **Évitez** les abréviations et écrivez clairement en lettres moulées • **Inscrivez** les achats en vrac	**Ce produit était-il . . . ?** **A** = frais **B** = congelé **C** = en boîte / embouteillé / en contenant de plastique **D** = séché **E** = autre **Inscrivez la lettre code**	**Acheté dans . . .** **A** = supermarché **B** = marché ou magasin d'alimentation spécialisé **C** = dépanneur **D** = autre **Inscrivez la lettre code**	**Nombre de . . .** conserves, bouteilles, cartons, sacs, boîtes, paquets, pommes, unités, douzaines, etc **Inscrivez la quantité**	**Poids ou volume par unité** Exemples 1,2 kg 2,5 litres 5,9 lb 19 oz unité douzaine paquet	**Coût total** **Exclure** toutes les taxes $ c	À l'usage du bureau
Exemple: *OIGNONS VERTS*	*A*	*B*	*1*	*PAQUET*	*79*	

Aliments et boissons achetés dans un RESTAURANT

Si vous n'avez rien acheté, cochez (✓) ici ☐

Repas ou collation de qui ? **Inscrivez :** • aliments de restaurants, cafétérias, mets à emporter, distributeurs, cantines, etc. • collations, boissons non alcoolisées, crème glacée, bonbons, etc. • repas achetés pour des invités **N'inscrivez pas** les boissons alcoolisées	**Type de repas :** **A** = Déjeuner **B** = Dîner **C** = Souper **D** = Collation et autres aliments **Inscrivez la lettre code**	**Nombre de repas achetés** **Inscrivez « 0 » si collations ou autre aliment**	**Type de restaurant :** **A** = Service aux tables **B** = Restauration rapide **C** = Cafétéria **D** = Autre **Inscrivez la lettre code**	**Coût total** **Inclure** les taxes et pourboires concernant les aliments et les boissons non alcoolisées **Exclure** les taxes et pourboires concernant les boissons alcoolisées $ c
Exemple: *DÎNER DU PÈRE ET DU FILS*	*B*	*2*	*B*	*8 69*

Si vous n'avez pas assez d'espace pour vos achats de la journée, utilisez les pages 20 à 23, « PAGE SUPPLÉMENTAIRE ».

7

Pour information seulement